Mathematics of Stochastic Processes & Randomness

Volume II

Stochastic Calculus & Differential Equations

Itô, Stratonovich, Fokker–Planck

Silvano Salvador

2026

Contents

Preface viii

I Foundations: Filtrations, Martingales, and Brownian Motion 1

1 Probability Spaces, Filtrations, and Adapted Processes 2
- 1.1 Review of Measure-Theoretic Probability . . . 2
- 1.2 Stopping Times . . . 3
- 1.3 Operations on Stopping Times . . . 3
- 1.4 Uniform Integrability . . . 4

2 Continuous-Time Martingales 6
- 2.1 Definition and Basic Properties . . . 6
- 2.2 Quadratic Variation and the Doob–Meyer Decomposition . . . 6
- 2.3 The Burkholder–Davis–Gundy Inequality . . . 7
- 2.4 Maximal Inequalities . . . 8
- 2.5 Local Martingales . . . 8

3 Brownian Motion: Construction and Path Properties 9
- 3.1 Definition and Characterisation . . . 9
- 3.2 Construction via the Wiener Measure . . . 10
- 3.3 Brownian Motion as a Gaussian Process . . . 10
- 3.4 Path Properties . . . 10
- 3.5 Symmetry Properties and the Reflection Principle . . . 11
- 3.6 The Law of the Iterated Logarithm . . . 12
- 3.7 The Markov Property and Blumenthal's Zero-One Law . . . 13

4 Multidimensional Brownian Motion and Brownian Filtrations 14
- 4.1 d-Dimensional Brownian Motion . . . 14
- 4.2 Recurrence and Transience . . . 14
- 4.3 Brownian Motion and Harmonic Functions . . . 15

4.4 The Natural Filtration and the Blumenthal–Getoor Augmentation . . . 15
4.5 Correlated Brownian Motions and the Cholesky Construction 15
4.6 Bessel Processes 16

5 Brownian Motion in Finance and Physics **18**
5.1 Brownian Motion as a Model of Diffusion 18
5.2 Geometric Brownian Motion and the Black–Scholes Model 18
5.3 Ornstein–Uhlenbeck Process 19
5.4 The Brownian Bridge 19
5.5 The Karhunen–Loève Expansion 20
5.6 Brownian Motion and Fractional Brownian Motion 21

II Itô Calculus **22**

6 The Itô Integral **23**
6.1 Motivation: Failure of the Riemann–Stieltjes Integral 23
6.2 The Space $\mathcal{H}^2$ 23
6.3 The Integral for Simple Processes 23
6.4 Extension by Density 24
6.5 The Cameron–Martin Theorem 25
6.6 The Itô Representation Theorem and Chaos Decomposition 26
6.7 Extension to Local Martingale Integrands 26
6.8 The Multidimensional Itô Integral 27

7 Itô's Formula **28**
7.1 The One-Dimensional Itô Formula 28
7.2 The General Itô Formula 29
7.3 The Doléans–Dade Exponential 29
7.4 Lévy's Characterisation Revisited 30

8 Semimartingales and the General Stochastic Integral **31**
8.1 Semimartingales 31
8.2 The Quadratic Covariation 31
8.3 The Predictable σ-Algebra and Predictable Processes 32
8.4 The Itô Product Rule for Semimartingales 32
8.5 The Doléans-Dade Formula and Special Semimartingales 33

9 The Girsanov Theorem and Change of Measure **34**
9.1 Absolute Continuity of Measures 34

9.2 Novikov's Condition and the Girsanov Density . . . 34
9.3 Girsanov's Theorem . . . 35
9.4 Multidimensional Girsanov . . . 35
9.5 Relative Entropy and the Entropic Cost of Measure Change . . . 36

10 The Martingale Representation Theorem **38**
10.1 Statement and Proof . . . 38
10.2 Consequences: Completeness of the Black–Scholes Market . . . 38
10.3 The Multidimensional Representation Theorem . . . 39
10.4 The Clark–Ocone Formula . . . 39

11 Stochastic Processes with Jumps: A Brief Overview **41**
11.1 Poisson Processes and Compound Poisson Processes . . . 41
11.2 The Poisson Stochastic Integral and Compensated Integral . . . 42

III Stochastic Differential Equations **44**

12 Existence and Uniqueness of SDEs **45**
12.1 Formulation . . . 45
12.2 Lipschitz Theory: Itô's Theorem . . . 45
12.3 Moment Estimates . . . 46
12.4 Continuity in Initial Conditions . . . 47

13 The Markov Property of SDE Solutions **48**
13.1 The Strong Markov Property . . . 48
13.2 The Generator of a Diffusion . . . 48
13.3 The Kolmogorov Equations and the Semigroup . . . 49

14 Stability and Lyapunov Methods for SDEs **51**
14.1 Notions of Stability . . . 51
14.2 Stochastic Lyapunov Functions . . . 51
14.3 Explosion and Non-explosion . . . 52
14.4 The Lyapunov Exponent and Almost-Sure Stability . . . 52

15 Numerical Methods for SDEs **54**
15.1 The Euler–Maruyama Scheme . . . 54
15.2 The Milstein Scheme . . . 54
15.3 The Stochastic Taylor Expansion . . . 55
15.4 Runge–Kutta Methods and Weak Convergence . . . 56

15.5 Stochastic Runge–Kutta and Implicit Methods 56

16 The Stochastic Flow and Malliavin Calculus (Introduction) 58
16.1 The Stochastic Flow of Diffeomorphisms 58
16.2 Introduction to Malliavin Calculus 58
16.3 The Chain Rule and the Skorokhod Integral 59
16.4 The Malliavin Covariance Matrix and Hörmander's Theorem 60

IV Fokker–Planck and Kolmogorov Equations 61

17 The Fokker–Planck Equation 62
17.1 Derivation from the SDE 62
17.2 Stationary Distributions and Detailed Balance 63
17.3 The Kramers–Smoluchowski Equation 63
17.4 The Probability Current and the Continuity Equation 63
17.5 Entropy Dissipation and Convergence to Equilibrium 64

18 The Backward Kolmogorov Equation and Feynman–Kac 66
18.1 The Backward Equation 66
18.2 Feynman–Kac Formula 66
18.3 Connection to the Heat Equation and Explicit Solutions 67
18.4 Feynman–Kac with a Source Term 67
18.5 The Killed Process Interpretation 68

19 Diffusion Processes and the Connection to Partial Differential Equations 70
19.1 Elliptic Equations and the Dirichlet Problem 70
19.2 Harmonic Functions and the Maximum Principle 70
19.3 Harnack Inequality and Regularity 71
19.4 Mean Exit Times and the Poisson Equation 71
19.5 The Principal Eigenvalue and Exponential Exit Asymptotics 71

20 Stochastic Optimal Control 73
20.1 The Stochastic Control Problem 73
20.2 The Hamilton–Jacobi–Bellman Equation 74
20.3 The Linear-Quadratic Regulator 74
20.4 Connection to Backward SDEs 75

V Advanced Topics **76**

21 Stratonovich Calculus **77**
21.1 Motivation: SDEs in Differential Geometry . . . 77
21.2 Conversion Between Itô and Stratonovich . . . 78
21.3 SDEs on Manifolds . . . 78
21.4 The Wong–Zakai Approximation Theorem . . . 78

22 Local Time and the Tanaka Formula **80**
22.1 Local Time of Brownian Motion . . . 80
22.2 The Occupation Times Formula and Generalised Itô–Tanaka . . . 80
22.3 Reflected and Absorbed Brownian Motion . . . 81
22.4 Joint Continuity and the Ray–Knight Theorems . . . 81
22.5 Itô's Excursion Theory . . . 82

23 Ergodic Theory for Diffusions **84**
23.1 Invariant Measures for SDEs . . . 84
23.2 Ergodicity and Mixing . . . 84
23.3 The Spectral Gap and Poincaré Inequality . . . 85
23.4 The Bakry–Émery Criterion . . . 85
23.5 Uniqueness of the Invariant Measure and Strong Feller Diffusions . . . 86

24 Rough Paths: An Introduction **88**
24.1 Motivation . . . 88
24.2 Young Integration and its Limits . . . 88
24.3 The Rough Path . . . 88
24.4 Brownian Motion as a Rough Path . . . 89
24.5 The Signature and Chen's Relation . . . 89
24.6 Rough Differential Equations and the Davie Expansion . . . 90

A Review of Measure Theory and Conditional Expectation **91**
A.1 Key Results . . . 91
A.2 Conditional Expectation . . . 91
A.3 The Borel–Cantelli Lemmas . . . 92
A.4 Versions and Regularity: The Kolmogorov Continuity Theorem . . . 92

B Gaussian Processes **94**
B.1 Gaussian Random Vectors . . . 94
B.2 Gaussian Hilbert Spaces . . . 94

B.3 Reproducing Kernel Hilbert Spaces 94
B.4 Wick Products and Hermite Polynomials 94
B.5 Conditioning of Gaussian Processes 95

C Notation and Symbol Table **97**

D Worked Applications **98**
D.1 European Option Pricing via Girsanov and the Black–Scholes Formula 98
D.2 The Ornstein–Uhlenbeck Process: Full Probabilistic Analysis 100
D.3 Barrier Option Pricing via the Reflection Principle 101
D.4 Overdamped Langevin Dynamics and the Convergence to Equilibrium 103
D.5 The Euler–Maruyama Scheme: Strong Convergence with Explicit Constants 105

Bibliography **107**

Preface

This second volume of *Mathematics of Stochastic Processes & Randomness* develops the theory of stochastic calculus and stochastic differential equations, building on the measure-theoretic and Markov-chain foundations laid in Volume I. The central objects here are continuous-time processes with sample paths of unbounded variation — above all, Brownian motion and its generalisations — and the calculus one can construct around them.

The subject originates in the work of Norbert Wiener (1923), who gave the first rigorous construction of Brownian motion as a probability measure on a function space, and of Kiyosi Itô (1944–1951), who defined the stochastic integral against Brownian motion and derived the change-of-variable formula that bears his name. The Fokker–Planck equation, relating the evolution of probability densities to the coefficients of an SDE, connects this analytic and probabilistic machinery to physics, where it describes the drift and diffusion of density under a stochastic force.

The volume is organised as follows. Part I (Chapters 1–5) lays the probabilistic groundwork: filtrations, martingales in continuous time, Brownian motion and its path properties. Part II (Chapters 6–11) develops the Itô integral and calculus, including Itô's formula, Girsanov's theorem, and the martingale representation theorem. Part III (Chapters 12–16) treats stochastic differential equations: existence and uniqueness, the Markov and strong Markov properties of solutions, stability, and numerical simulation. Part IV (Chapters 17–21) is devoted to the Fokker–Planck and Kolmogorov equations, connections to PDEs, and the theory of diffusion processes. Part V (Chapters 22–25) covers advanced topics: Stratonovich calculus, local time and the Tanaka formula, reflected and absorbed diffusions, and an introduction to rough paths. Appendices collect reference material on measure theory, Gaussian processes, and notation.

Prerequisites are a solid course in measure-theoretic probability (sigma-algebras, Lebesgue integration, conditional expectation) and familiarity with the material of Volume I, particularly continuous-time Markov chains and their generators. Functional analysis at the level of Hilbert spaces is used in several places and is recalled as needed.

Throughout, the emphasis is on rigour and on the interplay between the probabilistic and analytic viewpoints. Each chapter contains worked examples and concludes with exercises of varying difficulty. A curated bibliography points to the primary sources and to modern treatments for further reading.

Part I

Foundations: Filtrations, Martingales, and Brownian Motion

Chapter 1

Probability Spaces, Filtrations, and Adapted Processes

1.1 Review of Measure-Theoretic Probability

We work throughout on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Completeness means that every subset of a $\mathbb{P}$-null set belongs to $\mathcal{F}$; this technical requirement will be silently assumed to hold after completion wherever necessary.

Definition

A **stochastic process** (indexed by $[0,\infty)$) is a family $(X_t)_{t\ge 0}$ of random variables $X_t\colon \Omega \to \mathbb{R}^d$. The map $t \mapsto X_t(\omega)$ is the **sample path** or **trajectory** of ω.

For each ω the sample path is an element of $(\mathbb{R}^d)^{[0,\infty)}$. We frequently impose regularity conditions on these paths: continuity, right-continuity with left limits (càdlàg), or measurability in the product sense.

Definition: Filtration

A **filtration** is a non-decreasing family $(\mathcal{F}_t)_{t\ge 0}$ of sub-σ-algebras of $\mathcal{F}$: $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for all $0 \le s \le t$. The filtration is **right-continuous** if $\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s$ for all t.

Intuitively, $\mathcal{F}_t$ encodes the information available up to time t. The right-continuity condition is a standard regularity assumption; it holds, for instance, for the natural filtration of Brownian motion after augmentation with $\mathbb{P}$-null sets.

Definition: Adapted Process

A process $(X_t)_{t\ge 0}$ is **adapted** to $(\mathcal{F}_t)_{t\ge 0}$ if X_t is $\mathcal{F}_t$-measurable for every $t \ge 0$.

Definition: Progressively Measurable Process

A process $(X_t)_{t\geq 0}$ is **progressively measurable** with respect to $(\mathcal{F}_t)_{t\geq 0}$ if, for every $t \geq 0$, the map $(s,\omega) \mapsto X_s(\omega)$ from $([0,t]\times\Omega, \mathcal{B}([0,t])\otimes\mathcal{F}_t)$ to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is measurable.

Every adapted càdlàg process is progressively measurable. Progressive measurability is the correct hypothesis for the stochastic integral to be well-defined.

1.2 Stopping Times

Definition: Stopping Time

A random variable $\tau\colon \Omega \to [0,\infty]$ is a **stopping time** with respect to $(\mathcal{F}_t)_{t\geq 0}$ if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.

Example: First hitting time

Let (X_t) be adapted and continuous, and let $A \subseteq \mathbb{R}^d$ be a closed set. The **first hitting time** $\tau_A := \inf\{t \geq 0 : X_t \in A\}$ is a stopping time (with the convention $\inf\emptyset = \infty$). Proof: $\{\tau_A \leq t\} = \bigcup_{q\in\mathbb{Q}\cap[0,t]}\{X_q \in A\} \in \mathcal{F}_t$ when A is closed and paths are continuous.

Definition: Stopped σ-algebra

If τ is a stopping time, the **stopped σ-algebra** is

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0\}.$$

The class $\mathcal{F}_\tau$ captures "what is known at time τ". One verifies that $\mathcal{F}_\tau$ is indeed a σ-algebra, and that τ is $\mathcal{F}_\tau$-measurable.

1.3 Operations on Stopping Times

Stopping times are closed under several natural operations, all of which are used repeatedly throughout this volume whenever a localisation argument is required.

Stability Under Lattice and Sum Operations

If τ, σ are stopping times with respect to $(\mathcal{F}_t)_{t\geq 0}$, then so are:

(i) $\tau \wedge \sigma$ and $\tau \vee \sigma$ (minimum and maximum);

(ii) $\tau + \sigma$, provided σ is independent of $\mathcal{F}_\tau$ or, more generally, is itself measurable in a suitably shifted filtration;

(iii) $\tau_n \nearrow \tau$ if τ_n are stopping times increasing to τ (in fact $\{\tau \leq t\} = \bigcap_n \{\tau_n \leq$

$t\}$, an intersection of events in $\mathcal{F}_t$);

(iv) $\tau_n \searrow \tau$ if $(\mathcal{F}_t)_{t\geq 0}$ is right-continuous (this is precisely where right-continuity of the filtration is needed, since $\{\tau < t\} = \bigcup_n \{\tau_n < t\} \in \mathcal{F}_t$ gives τ measurable with respect to $\mathcal{F}_{t+} = \mathcal{F}_t$, not immediately with respect to $\mathcal{F}_t$ itself without right-continuity).

Remark: Why $\tau \wedge \sigma$ Matters in Practice

The operation $\tau \wedge \sigma$ is the basis of **localisation**, the technique used throughout Chapters 2, 6, and 12 to extend results from bounded or nice processes to general ones: given a process with a property that holds only locally, one constructs an increasing sequence of stopping times $\tau_n \nearrow \infty$ such that the stopped process $X_{t\wedge\tau_n}$ has the desired property for each n (e.g. is a true martingale, or has a bounded integrand), and then passes to the limit. The local martingale (Section 2.5) and the localised stochastic integral (Section 6.7) are both defined exactly via this mechanism.

Comparison of σ-Algebras

If $\tau \leq \sigma$ are stopping times, then $\mathcal{F}_\tau \subseteq \mathcal{F}_\sigma$. If $\tau = c$ is a deterministic constant, $\mathcal{F}_\tau = \mathcal{F}_c$, recovering the ordinary filtration σ-algebra as a special case of the stopped one.

Example: A Random Variable that is Not a Stopping Time

Let (B_t) be Brownian motion and let $L := \sup\{t \leq 1 : B_t = 0\}$, the **last** zero of B before time 1. Although L is a well-defined, finite random variable, it is *not* a stopping time: deciding whether $L \leq t$ for $t < 1$ requires knowledge of the path on the entire interval $[t, 1]$ (specifically, whether B avoids 0 on $(t, 1]$), which is not available from $\mathcal{F}_t$ alone. This distinction between hitting times (always stopping times, by the example above) and "last exit" type random times (typically not) is fundamental in excursion theory (Chapter 22) and explains why such times require an entirely different toolkit — they are examples of what is called a **co-terminal** or **honest** random time, not amenable to optional stopping arguments.

1.4 Uniform Integrability

Definition: Uniform Integrability

A family of random variables $\{X_\alpha\}_{\alpha\in I}$ is **uniformly integrable** (UI) if

$$\lim_{K\to\infty} \sup_{\alpha\in I} \mathbb{E}[|X_\alpha| \mathbf{1}_{\{|X_\alpha|>K\}}] = 0.$$

Uniform integrability is the key condition for passing limits through expectations in continuous-time martingale theory. Every family dominated in L^p for some $p > 1$ is UI; so is any family of the form $\{\mathbb{E}[Y \mid \mathcal{G}]\}_{\mathcal{G}}$ for a fixed integrable Y.

Theorem: Vitali Convergence

Let $\{X_n\}$ be a UI sequence and $X_n \to X$ in probability. Then $X \in L^1$ and $\mathbb{E}[|X_n - X|] \to 0$.

Chapter 2

Continuous-Time Martingales

2.1 Definition and Basic Properties

Definition: Continuous-Time Martingale

An adapted integrable process $(M_t)_{t\geq 0}$ is a **martingale** with respect to $(\mathcal{F}_t)_{t\geq 0}$ if

$$\mathbb{E}[M_t \mid \mathcal{F}_s] = M_s \quad \mathbb{P}\text{-a.s. for all } 0 \leq s \leq t.$$

If the equality is replaced by $\geq$ (resp. $\leq$) we obtain a **submartingale** (resp. **supermartingale**).

Example: L^2 martingales

Let $\xi \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and define $M_t := \mathbb{E}[\xi \mid \mathcal{F}_t]$. The tower property of conditional expectation implies (M_t) is a martingale. Choosing a càdlàg version (guaranteed by regularity theorems) gives a standard construction.

Theorem: Doob's Optional Stopping (continuous time)

Let (M_t) be a UI martingale and let $\sigma \leq \tau$ be bounded stopping times. Then $\mathbb{E}[M_\tau \mid \mathcal{F}_\sigma] = M_\sigma$ a.s.

Sketch. Approximate σ, τ by discrete stopping times taking values in $\{k/n : k \in \mathbb{N}_0\}$ and apply the discrete optional stopping theorem, then pass to the limit using UI and Vitali's theorem. □

2.2 Quadratic Variation and the Doob–Meyer Decomposition

For a square-integrable martingale (M_t), the **quadratic variation** $\langle M \rangle_t$ is the unique increasing predictable process such that $M_t^2 - \langle M \rangle_t$ is a martingale.

Definition: Quadratic Variation

For a continuous semimartingale X, the **quadratic variation** is

$$[X]_t := \lim_{\|\Pi\|\to 0} \sum_i (X_{t_{i+1}} - X_{t_i})^2$$

where the limit is in probability over partitions $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = t\}$ with mesh $\|\Pi\| \to 0$.

For a continuous local martingale, $[M]_t = \langle M \rangle_t$. For Brownian motion, we will see $[B]_t = t$ a.s.

Theorem: Doob–Meyer Decomposition

Every submartingale $(X_t)_{t\in[0,T]}$ with right-continuous paths that is of class (D) (UI family $\{X_\tau : \tau \text{ stopping time}\}$) admits a unique decomposition

$$X_t = M_t + A_t$$

where (M_t) is a uniformly integrable martingale and (A_t) is a natural increasing predictable process with $A_0 = 0$.

2.3 The Burkholder–Davis–Gundy Inequality

Doob's L^p inequality bounds the running maximum of a submartingale in terms of its terminal value. For stochastic integrals, the natural substitute is the **Burkholder–Davis–Gundy** (BDG) inequality, which relates the L^p norm of the running maximum to the L^p norm of the quadratic variation.

Theorem: Burkholder–Davis–Gundy Inequality

For every $p > 0$ there exist universal constants $0 < c_p \le C_p < \infty$ such that for every continuous local martingale (M_t) with $M_0 = 0$ and every stopping time τ,

$$c_p\, \mathbb{E}\big[[M]_\tau^{p/2}\big] \;\le\; \mathbb{E}\Big[\sup_{0\le t\le \tau} |M_t|^p\Big] \;\le\; C_p\, \mathbb{E}\big[[M]_\tau^{p/2}\big].$$

For $p = 2$ one may take $c_2 = 1$ and $C_2 = 4$.

Proof for $p = 2$ (upper bound). Let $M^* = \sup_{t\le\tau} |M_t|$. Since $M_t^2 - [M]_t$ is a local martingale, optional stopping (for a bounded τ) gives $\mathbb{E}[M_\tau^2] = \mathbb{E}[[M]_\tau]$. By Doob's L^2 maximal inequality applied to the submartingale $|M_t|^2$:

$$\mathbb{E}[(M^*)^2] \le 4\,\mathbb{E}[M_\tau^2] = 4\,\mathbb{E}[[M]_\tau],$$

which is the upper BDG bound with $C_2 = 4$. The lower bound $\mathbb{E}[(M^*)^2] \ge \mathbb{E}[M_\tau^2] = \mathbb{E}[[M]_\tau]$ follows from $(M^*)^2 \ge M_\tau^2$. □

Corollary: L^p control of the Itô integral

For $H \in \mathcal{H}^2_{\text{loc}}$ and $p \geq 2$, the stochastic integral $I_t = \int_0^t H_s \, \mathrm{d}W_s$ satisfies

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |I_t|^p\right] \leq C_p \, \mathbb{E}\left[\left(\int_0^T H_s^2 \, \mathrm{d}s\right)^{p/2}\right].$$

The BDG inequality is one of the most frequently cited tools in stochastic analysis: it appears in the proofs of strong convergence of numerical schemes (Chapter 14), the existence theory for SDEs (Chapter 12), and the continuity of stochastic integrals in the initial condition.

2.4 Maximal Inequalities

Doob's L^p Maximal Inequality

Let $(M_t)_{t\in[0,T]}$ be a non-negative submartingale with $\mathbb{E}[M_T^p] < \infty$ for some $p > 1$. Then

$$\mathbb{E}\left[\sup_{0\leq t\leq T} M_t^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[M_T^p].$$

This inequality is fundamental for controlling the supremum of the Itô integral; it will be applied repeatedly in Chapter 7 when establishing the L^2 theory.

2.5 Local Martingales

Definition: Local Martingale

An adapted process $(M_t)_{t\geq 0}$ is a **local martingale** if there exist stopping times $\tau_n \nearrow \infty$ a.s. such that each stopped process $M_t^{\tau_n} := M_{t\wedge\tau_n}$ is a uniformly integrable martingale.

Every L^2-bounded martingale is a local martingale; the converse fails. The stochastic integral defined in Chapter 6 will be a local martingale in general and a true martingale under an additional integrability condition.

Chapter 3

Brownian Motion: Construction and Path Properties

3.1 Definition and Characterisation

Definition: Brownian Motion

A **(standard) Brownian motion** (or **Wiener process**) is a stochastic process $(B_t)_{t\geq 0}$ such that:

(i) $B_0 = 0$ a.s.;

(ii) (B_t) has **independent increments**: for $0 \leq t_1 < t_2 < \cdots < t_n$, the increments $B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent;

(iii) **Stationary Gaussian increments**: $B_t - B_s \sim \mathcal{N}(0, t-s)$ for all $0 \leq s < t$;

(iv) The map $t \mapsto B_t(\omega)$ is continuous a.s.

Conditions (i)–(iii) determine all finite-dimensional marginals of (B_t) via the Kolmogorov extension theorem; condition (iv) selects a continuous modification.

Theorem: Lévy's Characterisation of Brownian Motion

A continuous adapted process $(M_t)_{t\geq 0}$ with $M_0 = 0$ is a Brownian motion if and only if:

(i) (M_t) is a local martingale;

(ii) $[M]_t = t$ for all $t \geq 0$ a.s.

This characterisation is extremely useful: it allows us to identify processes as Brownian motions without checking the distribution directly. It is also the cornerstone of the proof that the Itô integral against Brownian motion is itself a Brownian motion up to a time-change.

3.2 Construction via the Wiener Measure

The classical Wiener construction realises Brownian motion as a probability measure on $C([0,\infty))$. Define $\Omega = C([0,\infty),\mathbb{R})$ with the topology of uniform convergence on compact sets (giving the Borel σ-algebra $\mathcal{B}$), and let $X_t(\omega) = \omega(t)$ be the coordinate process. The **Wiener measure** $\mathbb{P}^W$ is the unique probability measure on $(\Omega,\mathcal{B})$ under which (X_t) satisfies (i)–(iii) of the above definition. Its existence follows from Kolmogorov's consistency theorem together with Kolmogorov's continuity criterion.

Theorem: Kolmogorov Continuity Criterion

Let $(X_t)_{t\in[0,T]}$ be a stochastic process such that

$$\mathbb{E}[|X_t - X_s|^{\alpha}] \le C\,|t-s|^{1+\beta}$$

for constants $\alpha,\beta,C>0$ and all $s,t\in[0,T]$. Then (X_t) has a modification with Hölder-continuous paths of exponent γ for every $\gamma\in(0,\beta/\alpha)$.

For Brownian motion, $\mathbb{E}[|B_t-B_s|^{2k}] = (2k-1)!!\,|t-s|^k$ for $k\in\mathbb{N}$, so taking $k=2$ gives $\alpha=4$, $\beta=1$, and paths are Hölder of order $\gamma<1/4$. The true Hölder regularity is any $\gamma<1/2$.

3.3 Brownian Motion as a Gaussian Process

The process $(B_t)_{t\ge0}$ is a Gaussian process with mean function $m(t)=0$ and covariance function $K(s,t)=\min(s,t)$. The finite-dimensional distributions are hence

$$(B_{t_1},\dots,B_{t_n}) \sim \mathcal{N}(0,\Sigma), \qquad \Sigma_{ij} = \min(t_i,t_j).$$

This covariance structure encodes the Markov property: the correlation between B_s and B_t (for $s<t$) equals $\sqrt{s/t}$, reflecting the fact that after time s the future depends only on the present position.

3.4 Path Properties

Theorem: Non-differentiability

Almost surely, the sample paths of Brownian motion are nowhere differentiable.

Sketch. If $B_t(\omega)$ were differentiable at some t_0, the derivative would be the limit of $(B_{t_0+h}-B_{t_0})/h$ as $h\to0$. A Borel–Cantelli argument shows that $\limsup_{h\to0}|B_{t_0+h}-B_{t_0}|/\sqrt{h}=\infty$ a.s. for every fixed t_0, and a covering argument extends this to all t_0 simultaneously. $\square$

Theorem: Unbounded Variation

Almost surely, Brownian motion has unbounded total variation on every interval $[s,t]$ with $s<t$.

Theorem: Quadratic Variation of Brownian Motion

For Brownian motion and any sequence of partitions $\Pi_n = \{0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t\}$ with $\|\Pi_n\| \to 0$,

$$\sum_{i=0}^{k_n-1} (B_{t_{i+1}^n} - B_{t_i^n})^2 \xrightarrow{L^2} t.$$

Hence $[B]_t = t$ a.s.

Proof. Let $S_n = \sum_i (B_{t_{i+1}^n} - B_{t_i^n})^2$. Then $\mathbb{E}[S_n] = \sum_i (t_{i+1}^n - t_i^n) = t$. For the variance,

$$\operatorname{Var}(S_n) = \sum_i \operatorname{Var}((B_{t_{i+1}^n} - B_{t_i^n})^2) = \sum_i 2(t_{i+1}^n - t_i^n)^2 \le 2t\|\Pi_n\| \to 0.$$

So $S_n \to t$ in L^2. □

The heuristic $(\mathrm{d}B_t)^2 = \mathrm{d}t$ (and $(\mathrm{d}B_t)(\mathrm{d}t) = 0$, $(\mathrm{d}t)^2 = 0$) is the essence of Itô's multiplication table.

3.5 Symmetry Properties and the Reflection Principle

Brownian motion enjoys a rich family of distributional symmetries:

(a) **Scaling**: $(c^{-1/2}B_{ct})_{t\ge 0} \overset{d}{=} (B_t)_{t\ge 0}$ for all $c>0$;

(b) **Reflection**: $(-B_t)_{t\ge 0} \overset{d}{=} (B_t)_{t\ge 0}$;

(c) **Time-reversal**: $(B_T - B_{T-t})_{0\le t\le T} \overset{d}{=} (B_t)_{0\le t\le T}$;

(d) **Time-inversion**: $(tB_{1/t})_{t>0} \overset{d}{=} (B_t)_{t>0}$ (with $B_0 = 0$).

Theorem: The Reflection Principle

Let $M_t = \max_{0\le s\le t} B_s$. For $a>0$ and $t>0$,

$$\mathbb{P}(M_t \ge a) = 2\mathbb{P}(B_t \ge a) = 2\left(1 - \Phi\left(\frac{a}{\sqrt{t}}\right)\right)$$

where Φ is the standard normal CDF.

Example: Distribution of the maximum

The joint density of (B_t, M_t) is

$$f_{B_t,M_t}(x,m) = \frac{2(2m-x)}{t\sqrt{2\pi t}} \exp\left(-\frac{(2m-x)^2}{2t}\right), \quad m \geq 0,\ x \leq m.$$

This formula, derived via the reflection principle, is the starting point for barrier option pricing in mathematical finance.

3.6 The Law of the Iterated Logarithm

The almost-sure fluctuation rate of Brownian motion is quantified by the **law of the iterated logarithm** (LIL), one of the deepest path-property results.

Theorem: Khintchine's Law of the Iterated Logarithm

For a standard Brownian motion $(B_t)_{t\geq 0}$,

$$\limsup_{t\to\infty} \frac{B_t}{\sqrt{2t\log\log t}} = 1 \quad \mathbb{P}\text{-a.s.}$$

and

$$\liminf_{t\to\infty} \frac{B_t}{\sqrt{2t\log\log t}} = -1 \quad \mathbb{P}\text{-a.s.}$$

Corollary: LIL at the Origin

By time-inversion symmetry $((tB_{1/t})_{t>0} \overset{d}{=} (B_t)_{t>0})$, the LIL also holds at the origin:

$$\limsup_{t\to 0^+} \frac{B_t}{\sqrt{2t\log\log(1/t)}} = 1 \quad \mathbb{P}\text{-a.s.}$$

The LIL shows that Brownian paths oscillate on a precise scale: the envelope $\pm\sqrt{2t\log\log t}$ is tight in the almost-sure sense. This is in sharp contrast with the L^2 scale $\sqrt{t}$ and the almost-sure Hölder scale t^γ for $\gamma < 1/2$ — each capturing a different aspect of Brownian regularity.

Example: LIL vs. CLT vs. Hölder

For Brownian motion:

(i) **CLT scale**: $B_t/\sqrt{t} \to \mathcal{N}(0,1)$ in distribution (no a.s. statement);

(ii) **Hölder scale**: $|B_t|/t^\gamma \to 0$ a.s. as $t \to \infty$, for any fixed $\gamma < 1/2$;

(iii) **LIL scale**: $B_t/\sqrt{2t\log\log t}$ oscillates between -1 and $+1$ a.s.

These three scales — $t^{1/2}$, $t^{1/2-\varepsilon}$, and $t^{1/2}(\log\log t)^{1/2}$ — are thus successively finer refinements of Brownian fluctuation, each capturing genuinely different information about the path.

Proposition: LIL and the Brownian Bridge

The functional LIL (Strassen's theorem) asserts that the sequence of rescaled paths $\{t \mapsto B_{nt}/\sqrt{2n\log\log n}\}_{n\geq 3}$ is relatively compact in $C([0,1])$, with the set of limit points being exactly the set $\mathcal{K}$ of absolutely continuous functions f on $[0,1]$ with $f(0) = 0$ and $\int_0^1 (f'(t))^2\,\mathrm{d}t \leq 1$. This functional version implies the scalar LIL by taking $t = 1$.

3.7 The Markov Property and Blumenthal's Zero-One Law

Theorem: Markov Property of Brownian Motion

Let τ be a stopping time with $\tau < \infty$ a.s. Then, conditionally on $\mathcal{F}_\tau$, the process $(\tilde{B}_t)_{t\geq 0} := (B_{\tau+t} - B_\tau)_{t\geq 0}$ is a standard Brownian motion independent of $\mathcal{F}_\tau$.

Theorem: Blumenthal's Zero-One Law

For Brownian motion with its natural augmented filtration, the tail σ-algebra $\mathcal{F}_{0+} = \bigcap_{t>0} \mathcal{F}_t$ is trivial: every $A \in \mathcal{F}_{0+}$ satisfies $\mathbb{P}(A) \in \{0,1\}$.

Corollary

The event $\{B_t > 0 \text{ for all sufficiently small } t > 0\}$ has probability 0, as does $\{B_t < 0 \text{ for all sufficiently small } t > 0\}$. Hence Brownian motion changes sign infinitely often in any neighbourhood of 0.

Chapter 4

Multidimensional Brownian Motion and Brownian Filtrations

4.1 d-Dimensional Brownian Motion

Definition

A **d-dimensional Brownian motion** is a process $\mathbf{B}_t = (B_t^1, \dots, B_t^d)$ where $B^1, \dots, B^d$ are independent one-dimensional Brownian motions. The covariance structure is $\mathbb{E}[B_t^i B_s^j] = \delta_{ij} \min(s,t)$.

The d-dimensional process has the same scaling, time-reversal, and Markov properties as the one-dimensional case. It is rotationally invariant: for any orthogonal matrix $O \in O(d)$, $(O\mathbf{B}_t)_{t\geq 0} \overset{d}{=} (\mathbf{B}_t)_{t\geq 0}$.

4.2 Recurrence and Transience

Theorem: Pólya–Kakutani Recurrence

(i) For $d \leq 2$, Brownian motion in $\mathbb{R}^d$ is recurrent: for every open set U, $\mathbb{P}(\mathbf{B}_t \in U \text{ for some } t \geq T) = 1$ for all T.

(ii) For $d \geq 3$, Brownian motion is transient: $|\mathbf{B}_t| \to \infty$ a.s. as $t \to \infty$.

The dichotomy $d \leq 2$ vs $d \geq 3$ reflects the classical distinction between recurrent and transient random walks (cf. Volume I, Chapter 17).

4.3 Brownian Motion and Harmonic Functions

Theorem: Kakutani's Solution to the Dirichlet Problem

Let $D \subset \mathbb{R}^d$ be a bounded domain with regular boundary ∂D, and let $f : \partial D \to \mathbb{R}$ be continuous. Define the **harmonic measure** solution

$$u(x) := \mathbb{E}^x[f(\mathbf{B}_{\tau_D})], \quad \tau_D = \inf\{t \geq 0 : \mathbf{B}_t \notin D\}.$$

Then u is the unique continuous function on $\overline{D}$ that is harmonic in D (i.e. $\Delta u = 0$) and equals f on ∂D.

This deep result connects Brownian motion to classical potential theory and will be vastly generalised in Part IV when we treat the Feynman–Kac formula.

4.4 The Natural Filtration and the Blumenthal–Getoor Augmentation

The **natural filtration** of Brownian motion is $\mathcal{F}_t^B = \sigma(B_s : 0 \leq s \leq t)$. After augmentation by the $\mathbb{P}$-null sets of $\mathcal{F} = \mathcal{F}_\infty^B$, the resulting filtration $(\bar{\mathcal{F}}_t^B)$ is right-continuous and complete (the *usual conditions* of stochastic analysis). All results in subsequent chapters assume this augmented filtration.

4.5 Correlated Brownian Motions and the Cholesky Construction

In many applications — multi-asset financial models, coupled physical systems, spatial diffusions with anisotropic noise — one needs Brownian motions with a prescribed instantaneous correlation rather than full independence. This is achieved by a linear transformation of independent components.

Definition: Correlated Brownian Motions

Let $B^1, \dots, B^d$ be independent standard Brownian motions and let $\Sigma = (\sigma_{ij})$ be a $d \times d$ symmetric positive semi-definite matrix with $\sigma_{ii} = 1$ for all i (a correlation matrix). Writing $\Sigma = LL^\top$ for the **Cholesky factorisation** (L lower triangular), define

$$W_t^i := \sum_{j=1}^{d} L_{ij} B_t^j, \qquad i = 1, \dots, d.$$

Then $(W_t^1, \dots, W_t^d)$ are standard Brownian motions with $\mathbb{E}[W_t^i W_t^j] = \sigma_{ij}\, t$ for all i, j.

Proof. Each $W_t^i = \sum_j L_{ij} B_t^j$ is a linear combination of independent Gaussians, hence

Gaussian, with mean zero and continuous paths inherited from the B^j. For the covariance structure, independence of increments gives

$$\mathbb{E}[W_t^i W_t^j] = \sum_{k,\ell} L_{ik} L_{j\ell}\, \mathbb{E}[B_t^k B_t^\ell] = t \sum_k L_{ik} L_{jk} = t\,(LL^\top)_{ij} = t\,\sigma_{ij}.$$

Taking $j = i$ shows $\mathbb{E}[(W_t^i)^2] = \sigma_{ii} t = t$, so each W^i has the variance of a standard Brownian motion; stationarity and independence of increments transfer directly from the B^j, and Lévy's characterisation (Chapter 9) confirms that each W^i is itself a standard Brownian motion. □

Example: Two Correlated Brownian Motions

For $d = 2$ with correlation $\rho \in (-1, 1)$, the correlation matrix is $\Sigma = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$, and the Cholesky factor is $L = \left(\begin{smallmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{smallmatrix}\right)$. Setting $W_t^1 = B_t^1$ and $W_t^2 = \rho B_t^1 + \sqrt{1-\rho^2}\, B_t^2$ produces two standard Brownian motions with $\mathbb{E}[W_t^1 W_t^2] = \rho t$. This construction underlies multi-asset Black–Scholes models, where ρ represents the instantaneous correlation between two stock returns.

4.6 Bessel Processes

The radial part of multidimensional Brownian motion gives rise to a one-dimensional diffusion family that generalises naturally to non-integer dimension and plays a central role in finance (the CIR interest-rate model) and in the theory of local times.

Definition: Bessel Process

For integer $d \geq 1$ and $\mathbf{B}_t$ a d-dimensional Brownian motion started at $\mathbf{B}_0 = x \neq 0$, the **Bessel process of dimension** d is $R_t := |\mathbf{B}_t|$. More generally, for any real parameter $\delta > 0$ (the **order** or **dimension parameter**), the squared Bessel process $X_t = R_t^2$ satisfies the SDE

$$\mathrm{d}X_t = \delta\, \mathrm{d}t + 2\sqrt{X_t}\, \mathrm{d}W_t, \qquad X_0 = x_0 \geq 0.$$

Itô's Formula for the Radial Part

For integer $d \geq 2$ and $\mathbf{B}_t \neq 0$, applying Itô's formula (Chapter 7) to $R_t = |\mathbf{B}_t| = (\sum_i (B_t^i)^2)^{1/2}$ gives

$$\mathrm{d}R_t = \mathrm{d}\beta_t + \frac{d-1}{2R_t}\, \mathrm{d}t,$$

where $\beta_t := \int_0^t \sum_i \frac{B_s^i}{R_s}\, \mathrm{d}B_s^i$ is a standard one-dimensional Brownian motion (by Lévy's characterisation, since $\mathrm{d}[\beta]_t = \sum_i \left(\frac{B_t^i}{R_t}\right)^2 \mathrm{d}t = \mathrm{d}t$).

Remark: The Drift Term and Dimension Dependence

The drift $\frac{d-1}{2R_t}$ is the discrete trace of the radial Laplacian in d dimensions, and it is precisely this term that drives the recurrence/transience dichotomy of Theorem 5.2 (Pólya–Kakutani): for $d = 1$ there is no drift and $R_t = |B_t|$ oscillates recurrently; for $d = 2$ the drift is positive but only barely repels the origin (logarithmically, consistent with neighbourhood recurrence in the plane); for $d \geq 3$ the drift $\frac{d-1}{2R_t}$ is strong enough near the origin to push R_t to infinity almost surely, matching transience.

Theorem: Feller's Boundary Classification

The squared Bessel process BESQ^δ with $\delta > 0$ started at $x_0 > 0$ satisfies:

(i) If $\delta \geq 2$, the process never hits 0 (0 is an **entrance boundary**, unreachable from the interior).

(ii) If $0 < \delta < 2$, the process hits 0 in finite time a.s. and, if $\delta > 0$, is instantaneously reflected (the origin is **regular**).

(iii) If $\delta = 0$, the origin is **absorbing**: once reached, the process remains at 0.

The squared Bessel process is the diffusion underlying the **Cox–Ingersoll–Ross (CIR) model** for short-term interest rates, $\mathrm{d}r_t = \kappa(\theta - r_t)\,\mathrm{d}t + \sigma\sqrt{r_t}\,\mathrm{d}W_t$, which is a time-changed and scaled BESQ process; the Feller condition $2\kappa\theta \geq \sigma^2$ for strict positivity of r_t is exactly the condition $\delta \geq 2$ of the theorem above, after matching parameters.

Chapter 5

Brownian Motion in Finance and Physics

5.1 Brownian Motion as a Model of Diffusion

In physics, diffusion is the phenomenon by which particles spread from regions of high concentration to low concentration. The classical model is the **heat equation**

$$\frac{\partial u}{\partial t} = \frac{1}{2}\Delta u,$$

where $u(t,x)$ represents the particle density at time t and position $x \in \mathbb{R}^d$. The connection to Brownian motion is the following: if $B_0 = x$, then the density of B_t is the **heat kernel**

$$p(t,x,y) = \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|y-x|^2}{2t}\right),$$

which satisfies $\partial_t p = \frac{1}{2}\Delta_y p$.

5.2 Geometric Brownian Motion and the Black–Scholes Model

Definition: Geometric Brownian Motion

A process $(S_t)_{t\geq 0}$ follows **geometric Brownian motion** (GBM) with drift $\mu \in \mathbb{R}$ and volatility $\sigma > 0$ if

$$S_t = S_0 \exp\left(\left(\mu - \tfrac{\sigma^2}{2}\right)t + \sigma B_t\right).$$

Equivalently, (S_t) satisfies the SDE $\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t$ (to be given precise meaning in Chapter 6).

The process (S_t) is the canonical model for asset prices in the Black–Scholes theory. Its use of stochastic calculus to derive the option pricing formula was the central

achievement of Black and Scholes (1973) and Merton (1973), work for which Scholes and Merton received the 1997 Nobel Memorial Prize in Economics.

5.3 Ornstein–Uhlenbeck Process

Definition: Ornstein–Uhlenbeck Process

The **Ornstein–Uhlenbeck** (OU) process satisfies

$$\mathrm{d}X_t = -\theta X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad X_0 = x_0,$$

with $\theta > 0$, $\sigma > 0$. Its explicit solution (derived in Chapter 9 using the Itô formula) is

$$X_t = x_0 e^{-\theta t} + \sigma \int_0^t e^{-\theta(t-s)}\,\mathrm{d}W_s.$$

The OU process is the unique stationary Gaussian Markov process in continuous time, and serves as a model for mean-reverting phenomena (interest rates, velocity in statistical mechanics).

5.4 The Brownian Bridge

Definition: Brownian Bridge

The **Brownian bridge** on $[0,1]$ is the process

$$X_t := B_t - tB_1, \qquad t \in [0,1],$$

i.e. Brownian motion conditioned to return to 0 at time 1. Equivalently, (X_t) is the unique continuous Gaussian process with $X_0 = X_1 = 0$, mean zero, and covariance $\mathbb{E}[X_s X_t] = s(1-t)$ for $s \le t$.

Proposition

The Brownian bridge satisfies the SDE

$$\mathrm{d}X_t = -\frac{X_t}{1-t}\,\mathrm{d}t + \mathrm{d}W_t, \qquad t \in [0,1),$$

which can be solved explicitly to recover $X_t = (1-t)\int_0^t \frac{\mathrm{d}W_s}{1-s}$, an alternative representation showing directly that $X_t \to 0$ a.s. as $t \to 1^-$.

Remark: Applications

The Brownian bridge arises as the limiting process in the Kolmogorov–Smirnov statistic for empirical distribution functions: if F_n is the empirical CDF of n i.i.d. samples from a continuous distribution F, then $\sqrt{n}\,(F_n(F^{-1}(t)) - t) \Rightarrow X_t$ on $[0,1]$, a fact known as Donsker's theorem for the empirical process. The bridge also appears as the scaling limit of a simple random walk conditioned to return to the origin, and as a building block in the Karhunen–Loève expansion below.

5.5 The Karhunen–Loève Expansion

Theorem: Karhunen–Loève Expansion of Brownian Motion

Brownian motion on $[0,1]$ admits the a.s. uniformly convergent series representation

$$B_t = \sum_{n=0}^{\infty} Z_n \, \phi_n(t), \qquad \phi_n(t) = \frac{\sqrt{2}\sin\!\big((n+\tfrac{1}{2})\pi t\big)}{(n+\tfrac{1}{2})\pi},$$

where $(Z_n)_{n\geq 0}$ are i.i.d. standard normal random variables and (ϕ_n) are the eigenfunctions of the covariance operator $K(s,t) = \min(s,t)$ on $L^2[0,1]$, with eigenvalues $\lambda_n = \big((n+\tfrac{1}{2})\pi\big)^{-2}$.

Sketch. The covariance kernel $K(s,t) = \min(s,t)$ defines a compact, self-adjoint, positive operator on $L^2[0,1]$ via $(K\phi)(t) = \int_0^1 \min(s,t)\phi(s)\,\mathrm{d}s$. Differentiating the eigenvalue equation $K\phi = \lambda\phi$ twice in t gives $-\phi'' = \lambda^{-1}\phi$ with boundary conditions $\phi(0) = 0$ (from $K(0,t) = 0$) and $\phi'(1) = 0$ (from differentiating once and setting $t = 1$), whose solutions are exactly the ϕ_n above with eigenvalues λ_n. Mercer's theorem then gives $K(s,t) = \sum_n \lambda_n \phi_n(s)\phi_n(t)$, and setting $Z_n := \lambda_n^{-1/2}\int_0^1 \phi_n(t)\,\mathrm{d}W_t$ (a standard Gaussian by the Itô isometry, to be developed in Chapter 6) yields the stated expansion, which converges in $L^2(\Omega)$ uniformly in t by Mercer's theorem and a martingale argument. $\square$

Remark: Computational and Theoretical Uses

The Karhunen–Loève expansion gives a constructive, finite-dimensional approximation to Brownian paths — truncating the series at N terms produces the best L^2-approximation to B by a rank-N Gaussian process, which is the basis of spectral methods for simulating Brownian motion and for dimension reduction in stochastic PDEs. It also gives an alternative proof of the Itô–Nisio theorem on a.s. convergence of orthogonal series representations of Gaussian processes, and underlies the Cameron–Martin space construction revisited in Chapter 6.

5.6 Brownian Motion and Fractional Brownian Motion

A natural question is whether the independent-increments and Markov structure of Brownian motion can be relaxed while retaining self-similarity and Gaussianity. The answer is fractional Brownian motion, which introduces long-range dependence governed by a single parameter.

Definition: Fractional Brownian Motion

For $H \in (0,1)$ (the **Hurst parameter**), a **fractional Brownian motion** (fBm) is a centred Gaussian process $(B_t^H)_{t\geq 0}$ with $B_0^H = 0$ and covariance

$$\mathbb{E}[B_t^H B_s^H] = \tfrac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right).$$

The case $H = \frac{1}{2}$ recovers standard Brownian motion, since the formula collapses to $\min(s,t)$.

Self-Similarity and Increment Correlation

fBm is H-self-similar: $(B_{ct}^H)_{t\geq 0} \overset{d}{=} (c^H B_t^H)_{t\geq 0}$ for $c > 0$. Its increments are stationary but, unlike standard Brownian motion, *not* independent for $H \neq \frac{1}{2}$: for $s,t > 0$,

$$\mathbb{E}[(B_{t+s}^H - B_t^H)(B_t^H - B_{t-s}^H)] = \tfrac{1}{2}\left[(2s)^{2H} - 2s^{2H}\right] s^{2H-2H} \cdot s^{2H}$$

simplifies (after substitution $u = s$) to a quantity with sign $\operatorname{sgn}(H - \frac{1}{2})$: increments are **positively correlated** (persistent) for $H > \frac{1}{2}$ and **negatively correlated** (anti-persistent) for $H < \frac{1}{2}$.

Remark: Path Regularity and the Failure of Semimartingale Structure

By a Kolmogorov-continuity argument identical in spirit to Theorem 3.10, fBm has a version with Hölder-continuous paths of any exponent $< H$; in particular, for $H > \frac{1}{2}$ paths are *smoother* than Brownian motion, and for $H < \frac{1}{2}$ they are *rougher*. Crucially, for $H \neq \frac{1}{2}$, fBm is **not a semimartingale** (Chapter 8) and has unbounded quadratic variation behaviour: the p-variation is finite exactly for $p > 1/H$. This obstructs a direct Itô theory and motivates either Young integration (valid for $H > \frac{1}{2}$, Chapter 25) or the rough path theory developed there for $H < \frac{1}{2}$. fBm with $H \approx 0.1$–0.2 has been proposed as a model for rough volatility in financial markets, reviving interest in non-Markovian stochastic calculus.

Part II

Itô Calculus

Chapter 6

The Itô Integral

6.1 Motivation: Failure of the Riemann–Stieltjes Integral

The Riemann–Stieltjes integral $\int_0^t f(s)\,\mathrm{d}g(s)$ is well-defined whenever f is continuous and g has bounded variation. Since Brownian motion has unbounded variation on every interval, the classical Stieltjes theory fails for integrals $\int_0^t H_s\,\mathrm{d}B_s$.

The resolution, due to Itô (1944), is to build the integral via an L^2 isometry, exploiting the fact that Brownian motion has finite quadratic variation $[B]_t = t$.

6.2 The Space $\mathcal{H}^2$

Definition: The space $\mathcal{H}^2$

Let $T > 0$. The space $\mathcal{H}^2 = \mathcal{H}^2([0,T])$ consists of all progressively measurable processes $(H_t)_{t\in[0,T]}$ such that

$$\|H\|_{\mathcal{H}^2}^2 := \mathbb{E}\left[\int_0^T H_s^2\,\mathrm{d}s\right] < \infty.$$

Equipped with this norm (identifying processes that agree $\mathrm{d}t\otimes\mathrm{d}\mathbb{P}$-a.e.), $\mathcal{H}^2$ is a Hilbert space.

6.3 The Integral for Simple Processes

A process $H \in \mathcal{H}^2$ is **simple** (or **elementary**) if there exists a partition $0 = t_0 < t_1 < \cdots < t_n = T$ and bounded $\mathcal{F}_{t_i}$-measurable random variables ξ_i such that

$$H_t = \sum_{i=0}^{n-1} \xi_i \mathbf{1}_{(t_i, t_{i+1}]}(t).$$

For such H, the Itô integral is defined as the left-endpoint Riemann sum:

$$I(H)_T := \int_0^T H_s \, \mathrm{d}W_s := \sum_{i=0}^{n-1} \xi_i (B_{t_{i+1}} - B_{t_i}).$$

Theorem: Itô Isometry for Simple Processes

For a simple process H,

$$\mathbb{E}[I(H)_T^2] = \mathbb{E}\left[\int_0^T H_s^2 \, \mathrm{d}s\right].$$

Equivalently, $\|I(H)\|_{L^2} = \|H\|_{\mathcal{H}^2}$.

Proof. Expand $I(H)_T^2 = \sum_{i,j} \xi_i (B_{t_{i+1}} - B_{t_i}) \xi_j (B_{t_{j+1}} - B_{t_j})$. For $i < j$, ξ_j and $B_{t_{i+1}} - B_{t_i}$ are both $\mathcal{F}_{t_j}$-measurable (actually the latter is $\mathcal{F}_{t_{i+1}}$-measurable), and $B_{t_{j+1}} - B_{t_j}$ is independent of $\mathcal{F}_{t_j}$ with mean zero. So the cross terms vanish, and the diagonal terms give $\mathbb{E}[\xi_i^2 (t_{i+1} - t_i)]$. □

6.4 Extension by Density

Simple processes are dense in $\mathcal{H}^2$: for any $H \in \mathcal{H}^2$ there exist simple $H^n \to H$ in $\mathcal{H}^2$-norm. The isometry $\|I(H^n) - I(H^m)\|_{L^2} = \|H^n - H^m\|_{\mathcal{H}^2} \to 0$ shows $(I(H^n)_T)$ is Cauchy in $L^2(\mathbb{P})$; define

$$\int_0^T H_s \, \mathrm{d}W_s := \lim_{n \to \infty} I(H^n)_T \quad \text{in } L^2(\mathbb{P}).$$

The limit is independent of the approximating sequence.

Theorem: Properties of the Itô Integral

Let $H, K \in \mathcal{H}^2$ and $\alpha, \beta \in \mathbb{R}$. The following hold:

(i) **Linearity**: $\int_0^T (\alpha H_s + \beta K_s) \, \mathrm{d}W_s = \alpha \int_0^T H_s \, \mathrm{d}W_s + \beta \int_0^T K_s \, \mathrm{d}W_s$;

(ii) **Itô isometry**: $\mathbb{E}\left[\left(\int_0^T H_s \, \mathrm{d}W_s\right)^2\right] = \mathbb{E}\left[\int_0^T H_s^2 \, \mathrm{d}s\right]$;

(iii) **Martingale property**: The process $\left(\int_0^t H_s \, \mathrm{d}W_s\right)_{t \in [0,T]}$ is a continuous L^2-martingale;

(iv) **Zero mean**: $\mathbb{E}\left[\int_0^T H_s \, \mathrm{d}W_s\right] = 0$.

6.5 The Cameron–Martin Theorem

A fundamental result connecting the Wiener measure to its perturbations is the **Cameron–Martin theorem**, which characterises which shifts of Brownian motion paths are absolutely continuous with respect to the Wiener measure.

Definition: Cameron–Martin Space

The **Cameron–Martin space** $\mathcal{H}$ consists of functions $h : [0,T] \to \mathbb{R}$ of the form $h(t) = \int_0^t \dot{h}(s)\,\mathrm{d}s$ with $\dot{h} \in L^2([0,T])$ (i.e., h is absolutely continuous with square-integrable derivative). The Cameron–Martin norm is $\|h\|_{\mathcal{H}}^2 = \int_0^T \dot{h}(s)^2\,\mathrm{d}s$.

Theorem: Cameron–Martin

Under the Wiener measure $\mathbb{P}^W$, the shifted Brownian motion $(B_t + h(t))_{t\in[0,T]}$ has a distribution absolutely continuous with respect to $\mathbb{P}^W$ if and only if $h \in \mathcal{H}$. In that case, the Radon–Nikodym derivative is

$$\frac{\mathrm{d}\mathbb{P}^{W+h}}{\mathrm{d}\mathbb{P}^W} = \exp\left(\int_0^T \dot{h}(s)\,\mathrm{d}B_s - \tfrac{1}{2}\int_0^T \dot{h}(s)^2\,\mathrm{d}s\right) = \mathcal{E}\left(\int_0^{\cdot} \dot{h}(s)\,\mathrm{d}B_s\right)_T.$$

Proof. The result is a special case of Girsanov's theorem (Chapter 10) with constant (non-random) drift $\theta_s = \dot{h}(s)$. Since $h \in \mathcal{H}$ implies $\int_0^T \dot{h}(s)^2\,\mathrm{d}s < \infty$, Novikov's condition is satisfied and the Doléans exponential is a true martingale. □

Corollary: Shifts of non-Cameron–Martin type are singular

If $h \notin \mathcal{H}$ (e.g., $h(t) = t^{1/3}$ since $\dot{h}(t) = t^{-2/3}/3 \notin L^2([0,1])$), then the shifted measure $\mathbb{P}^{W+h}$ is singular with respect to $\mathbb{P}^W$: the two measures are mutually singular and no Radon–Nikodym derivative exists.

The Cameron–Martin theorem has a beautiful geometric interpretation: $\mathcal{H}$ is the tangent space to the Wiener measure, and only shifts in the "infinitely thin" subspace $\mathcal{H} \subset C([0,T])$ can be detected absolutely continuously. This is the probabilistic analogue of the fact that in infinite-dimensional spaces, "balls" and "cubes" of different sizes are orthogonal — a fundamental departure from finite-dimensional intuition.

Example: Application to sampling and MCMC on function spaces

In Bayesian nonparametric statistics, one often places a Gaussian process prior μ_0 on a function u (the unknown) and observes data $y = \mathcal{G}(u) + \eta$. By the Cameron–Martin theorem, the posterior measure μ_y is absolutely continuous with respect to μ_0 if and only if the data operator $\mathcal{G}$ maps into the Cameron–Martin space, which is a regularity condition on the observation model. MCMC methods on function spaces (e.g. the preconditioned Crank–Nicolson algorithm) are designed to propose moves within $\mathcal{H}$ to maintain absolute continuity.

6.6 The Itô Representation Theorem and Chaos Decomposition

The Itô representation theorem (Chapter 10) states that every square-integrable $\mathcal{F}_T$-measurable functional can be written as a stochastic integral. The **Wiener chaos decomposition** gives a more refined orthogonal decomposition.

Definition: Wiener Chaos

The n-th **Wiener chaos** $\mathcal{H}_n$ is the closed subspace of $L^2(\Omega)$ spanned by Hermite polynomials of the Gaussian random variables $\{I(h) = \int_0^T h(t)\,\mathrm{d}B_t : h \in L^2([0,T])\}$:

$$\mathcal{H}_n = \overline{\operatorname{span}}\big\{ H_n(I(h)/\,\|h\|_{L^2}) : h \in L^2([0,T]),\, \|h\|_{L^2} = 1 \big\},$$

where $H_n(x) = (-1)^n e^{x^2/2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2/2}$ are the Hermite polynomials.

Theorem: Wiener Chaos Decomposition

Every $F \in L^2(\Omega, \mathcal{F}_T, \mathbb{P})$ has a unique orthogonal decomposition

$$F = \mathbb{E}[F] + \sum_{n=1}^{\infty} I_n(f_n),$$

where $I_n(f_n) = \int_{[0,T]^n} f_n(t_1,\dots,t_n)\,\mathrm{d}B_{t_1}\cdots\mathrm{d}B_{t_n}$ is the n-fold Wiener integral of a symmetric $f_n \in L^2([0,T]^n)$, and

$$\mathbb{E}[F^2] = \mathbb{E}[F]^2 + \sum_{n=1}^{\infty} n!\;\|f_n\|^2_{L^2([0,T]^n)}\,.$$

The Wiener chaos decomposition is the probabilistic analogue of the Fourier expansion; it plays a central role in Malliavin calculus (Chapter 15), where $D_s F = \sum_{n\ge 1} n\, I_{n-1}(f_n(\cdot, s))$.

6.7 Extension to Local Martingale Integrands

The construction above requires $H \in \mathcal{H}^2$. For the theory of SDEs, one needs $\int_0^t H_s\,\mathrm{d}W_s$ to be defined for a larger class of processes. The extension goes as follows.

Definition: Local square-integrability

Let $\mathcal{H}^2_{\mathrm{loc}}$ denote the class of progressively measurable (H_t) with $\int_0^T H_s^2\,\mathrm{d}s < \infty$ a.s. for every $T > 0$.

For $H \in \mathcal{H}^2_{\mathrm{loc}}$, define the localising stopping times $\tau_n := \inf\{t : \int_0^t H_s^2\,\mathrm{d}s \ge n\} \wedge n$. Each stopped integrand $H\mathbf{1}_{[0,\tau_n]}$ lies in $\mathcal{H}^2$, and the local integrals are consistent. The

resulting process $\int_0^t H_s \, \mathrm{d}W_s$ is a continuous local martingale.

6.8 The Multidimensional Itô Integral

Let $\mathbf{B} = (B^1, \ldots, B^m)$ be an m-dimensional Brownian motion. The covariation $[B^i, B^j]_t = \delta_{ij} t$ reflects the independence of the components. For a progressively measurable $\mathbb{R}^{n \times m}$-valued process $H = (H^{ij})$ with $\sum_{i,j} \int_0^T (H_s^{ij})^2 \, \mathrm{d}s < \infty$ a.s., define

$$\left(\int_0^t H_s \, \mathrm{d}\mathbf{B}_s \right)^i = \sum_{j=1}^m \int_0^t H_s^{ij} \, \mathrm{d}B_s^j.$$

The quadratic variation of the i-th component is $\int_0^t \sum_j (H_s^{ij})^2 \, \mathrm{d}s$.

Chapter 7

Itô's Formula

7.1 The One-Dimensional Itô Formula

Theorem: Itô's Formula (one dimension)

Let $f:\mathbb{R}\to\mathbb{R}$ be twice continuously differentiable ($f\in C^2$), and let (X_t) be a continuous semimartingale of the form

$$X_t = X_0 + \int_0^t b_s\,\mathrm{d}s + \int_0^t \sigma_s\,\mathrm{d}W_s.$$

Then

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_s)\sigma_s^2\,\mathrm{d}s.$$

Equivalently, in differential notation:

$$\mathrm{d}f(X_t) = f'(X_t)\,\mathrm{d}X_t + \tfrac{1}{2}f''(X_t)\sigma_t^2\,\mathrm{d}t.$$

The extra term $\frac{1}{2}f''\sigma^2\,\mathrm{d}t$ is the *Itô correction*, absent from the classical chain rule. It arises precisely because $(\mathrm{d}X_t)^2 = \sigma_t^2\,\mathrm{d}t \neq 0$.

Sketch. By Taylor expansion,

$$f(X_{t_{i+1}}) - f(X_{t_i}) = f'(X_{t_i})(X_{t_{i+1}}-X_{t_i}) + \tfrac{1}{2}f''(X_{t_i})(X_{t_{i+1}}-X_{t_i})^2 + R_i$$

where $R_i = o((X_{t_{i+1}}-X_{t_i})^2)$ in probability. Summing over a partition, the first-order terms converge to $\int f'(X_s)\,\mathrm{d}X_s$ (standard stochastic integral), the second-order terms converge to $\frac{1}{2}\int f''(X_s)\sigma_s^2\,\mathrm{d}s$ (via the quadratic variation), and the remainder vanishes. □

Example: Exponential martingale

Let $X_t = \sigma B_t - \frac{1}{2}\sigma^2 t$ and $Z_t = e^{X_t}$. Applying Itô's formula with $f(x)=e^x$:

$$\mathrm{d}Z_t = e^{X_t}\,\mathrm{d}X_t + \tfrac{1}{2}e^{X_t}\sigma^2\,\mathrm{d}t = Z_t(\sigma\,\mathrm{d}W_t - \tfrac{1}{2}\sigma^2\,\mathrm{d}t) + \tfrac{1}{2}\sigma^2 Z_t\,\mathrm{d}t = \sigma Z_t\,\mathrm{d}W_t.$$

Since there is no $\mathrm{d}t$ term, (Z_t) is a local martingale. When σ is bounded, it is a

true martingale: the **stochastic exponential** or Doléans exponential.

Example: Quadratic variation via Itô

Applying Itô's formula to $f(x) = x^2$ and $X_t = B_t$:

$$B_t^2 = 2\int_0^t B_s \, \mathrm{d}W_s + t.$$

This recovers $[B]_t = t$ and shows $B_t^2 - t$ is a martingale.

7.2 The General Itô Formula

Theorem: Itô's Formula (general)

Let $f : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ be $C^{1,2}$ (once in t, twice in x), and let $\mathbf{X}_t = (X_t^1, \dots, X_t^d)$ be a continuous semimartingale with

$$\mathrm{d}X_t^i = b_t^i \, \mathrm{d}t + \sum_{j=1}^{m} \sigma_t^{ij} \, \mathrm{d}B_t^j.$$

Then

$$\mathrm{d}f(t, \mathbf{X}_t) = \partial_t f \, \mathrm{d}t + \sum_i \partial_{x_i} f \, \mathrm{d}X_t^i + \tfrac{1}{2} \sum_{i,k} \partial^2_{x_i x_k} f \; \mathrm{d}[X^i, X^k]_t$$

where $\mathrm{d}[X^i, X^k]_t = \sum_j \sigma_t^{ij} \sigma_t^{kj} \, \mathrm{d}t$.

Example: Product rule (integration by parts)

Taking $f(x, y) = xy$, $\mathbf{X}_t = (X_t, Y_t)$ two semimartingales:

$$\mathrm{d}(X_t Y_t) = Y_t \, \mathrm{d}X_t + X_t \, \mathrm{d}Y_t + \mathrm{d}[X, Y]_t.$$

This is the Itô product rule; it reduces to the classical formula when one process is of bounded variation (so $[X, Y] = 0$).

7.3 The Doléans–Dade Exponential

Definition: Stochastic Exponential

For a continuous semimartingale X with $X_0 = 0$, the **stochastic exponential** (or Doléans–Dade exponential) is

$$\mathcal{E}(X)_t := \exp\!\left(X_t - \tfrac{1}{2}[X]_t\right).$$

Theorem

$(\mathcal{E}(X)_t)$ is the unique continuous semimartingale Z satisfying $\mathrm{d}Z_t = Z_t\,\mathrm{d}X_t$ with $Z_0 = 1$.

The stochastic exponential plays a central role in the theory of equivalent martingale measures (Girsanov's theorem, Chapter 10).

7.4 Lévy's Characterisation Revisited

We can now give a proof of Lévy's characterisation theorem (stated in Chapter 3).

Proof of Lévy's theorem. Let (M_t) be a continuous local martingale with $[M]_t = t$. For fixed $\lambda \in \mathbb{R}$, apply Itô's formula to $e^{i\lambda M_t + \lambda^2 t/2}$:

$$\mathrm{d}e^{i\lambda M_t + \lambda^2 t/2} = e^{i\lambda M_t + \lambda^2 t/2}\big(i\lambda\,\mathrm{d}M_t + \tfrac{1}{2}\lambda^2\,\mathrm{d}t - \tfrac{1}{2}\lambda^2\,\mathrm{d}[M]_t\big) = i\lambda e^{i\lambda M_t + \lambda^2 t/2}\,\mathrm{d}M_t.$$

Since $[M]_t = t$, the $\mathrm{d}t$ terms cancel. The process is thus a local martingale; boundedness arguments (for bounded λ) promote it to a true martingale. Evaluating $\mathbb{E}[e^{i\lambda M_t} \mid \mathcal{F}_s] = e^{-\lambda^2(t-s)/2} e^{i\lambda M_s}$ and reading off the characteristic function identifies the increments as Gaussian. $\square$

Chapter 8

Semimartingales and the General Stochastic Integral

8.1 Semimartingales

Definition: Semimartingale

A **semimartingale** is an adapted càdlàg process X that can be written as $X_t = X_0 + M_t + A_t$ where (M_t) is a local martingale with $M_0 = 0$ and (A_t) is a finite-variation process with $A_0 = 0$.

The class of semimartingales is the natural domain of the stochastic integral: by the Bichteler–Dellacherie theorem, it is the largest class of integrators for which the integral behaves well (i.e. for which the map $H \mapsto \int H \, \mathrm{d}X$ is continuous in an appropriate sense).

8.2 The Quadratic Covariation

Definition: Quadratic Covariation

For two semimartingales X, Y, the **quadratic covariation** (or **bracket**) is

$$[X,Y]_t = XY_t - X_0Y_0 - \int_0^t X_{s-}\,\mathrm{d}Y_s - \int_0^t Y_{s-}\,\mathrm{d}X_s,$$

equivalently characterised as the unique finite-variation process such that $XY - [X,Y]$ is a local martingale, or as the limit of $\sum_i (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i})$ over refining partitions.

Key properties: $[X,X] = [X]$ (the quadratic variation); $[B^i, B^j] = \delta_{ij} t$; $[X, A] = 0$ when A has bounded variation.

8.3 The Predictable σ-Algebra and Predictable Processes

Definition: Predictable σ-algebra

The **predictable σ-algebra** $\mathcal{P}$ on $[0,\infty)\times\Omega$ is generated by all left-continuous adapted processes. A process is **predictable** if it is $\mathcal{P}$-measurable.

Every continuous adapted process is predictable. The predictable σ-algebra is strictly smaller than the progressive σ-algebra; it is the correct integrability condition for the integrand in stochastic integrals with respect to general semimartingales.

8.4 The Itô Product Rule for Semimartingales

Theorem: Integration by Parts

For semimartingales X, Y,

$$X_tY_t = X_0Y_0 + \int_0^t X_{s-}\,\mathrm{d}Y_s + \int_0^t Y_{s-}\,\mathrm{d}X_s + [X,Y]_t.$$

This is simply a rearrangement of the defining identity for $[X,Y]$ in the previous section, but its importance is best appreciated through special cases. When X, Y are continuous, $[X,Y]_t$ reduces to the familiar covariation from Chapter 2, and when one of them has finite variation, $[X,Y]\equiv 0$ and the formula becomes ordinary integration by parts for Riemann–Stieltjes integrals. When both X and Y are continuous semimartingales with $\mathrm{d}X = b_X\,\mathrm{d}t + \sigma_X\,\mathrm{d}W$ and $\mathrm{d}Y = b_Y\,\mathrm{d}t + \sigma_Y\,\mathrm{d}W$ (driven by the same Brownian motion), the formula recovers the Itô product rule used repeatedly in Chapters 9 and 12: $\mathrm{d}(X_tY_t) = X_t\,\mathrm{d}Y_t + Y_t\,\mathrm{d}X_t + \sigma_X\sigma_Y\,\mathrm{d}t$.

Example: Product of Two Geometric Brownian Motions

Let S^1_t, S^2_t be two correlated GBMs (Chapter 5) with $\mathrm{d}S^i_t = \mu_i S^i_t\,\mathrm{d}t + \sigma_i S^i_t\,\mathrm{d}W^i_t$ and $\mathrm{d}[W^1,W^2]_t = \rho\,\mathrm{d}t$. By the product rule,

$$\mathrm{d}(S^1_tS^2_t) = S^1_t\,\mathrm{d}S^2_t + S^2_t\,\mathrm{d}S^1_t + \sigma_1\sigma_2\rho\,S^1_tS^2_t\,\mathrm{d}t,$$

so $Z_t := S^1_tS^2_t$ is itself a geometric Brownian motion (after identifying the diffusion coefficient) with drift $\mu_1+\mu_2+\rho\sigma_1\sigma_2$ — a fact used directly in pricing options on a product of two correlated assets, such as a quanto option.

8.5 The Doléans-Dade Formula and Special Semimartingales

Definition: Special Semimartingale

A semimartingale $X = X_0 + M + A$ is **special** if the finite-variation part A can be chosen *predictable*. The decomposition $X = X_0 + M + A$ with A predictable is then unique, and is called the **canonical decomposition**.

Remark: Why Predictability Matters

An arbitrary semimartingale may admit multiple decompositions into local martingale plus finite-variation process; uniqueness is restored precisely by requiring the finite-variation part to be predictable. Every continuous semimartingale is automatically special, since the finite-variation part of a continuous process is automatically predictable (continuity implies left-continuity, the defining property of the generators of $\mathcal{P}$). The distinction becomes essential only for processes with jumps, such as the jump-diffusions of Chapter 10, where the compensator of the jump measure plays the role of the predictable drift.

Stability of the Semimartingale Class

The class of semimartingales is closed under the following operations, all essential for the theory developed in later chapters:

(i) C^2 **functions**: if X is a semimartingale and $f \in C^2$, then $f(X)$ is a semimartingale (this is the content of the general Itô formula, Chapter 7).

(ii) **Stochastic integration**: if X is a semimartingale and H is predictable and locally bounded, then $\int H \, \mathrm{d}X$ is a semimartingale.

(iii) **Absolutely continuous change of measure**: if X is a $\mathbb{P}$-semimartingale and $\mathbb{Q} \sim \mathbb{P}$, then X is also a $\mathbb{Q}$-semimartingale (though its canonical decomposition changes — this is precisely the content of the Girsanov theorem, our next chapter).

(iv) **Time change**: if τ_t is a (sufficiently regular) family of stopping times, X_{τ_t} is again a semimartingale.

This closure under change of measure is what makes semimartingales — rather than the more restrictive class of Itô processes — the natural setting for the Girsanov theory developed next: the property of *being* a semimartingale is invariant under equivalent change of measure, even though the specific drift and martingale parts are not.

Chapter 9

The Girsanov Theorem and Change of Measure

9.1 Absolute Continuity of Measures

Definition

Two probability measures $\mathbb{P}$ and $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ are **equivalent** (written $\mathbb{P} \sim \mathbb{Q}$) if they have the same null sets. The Radon–Nikodym derivative $\mathrm{d}\mathbb{Q}/\,\mathrm{d}\mathbb{P}$ is then a strictly positive random variable.

In finance and in the theory of SDEs, one often needs to change the underlying probability measure to simplify the drift of a process, transforming a process with drift into a Brownian motion under a new measure.

9.2 Novikov's Condition and the Girsanov Density

Definition: Girsanov density

Let $(\theta_t)_{t\in[0,T]}$ be a progressively measurable process. Define the **Girsanov density**

$$Z_T = \mathcal{E}\left(-\int_0^{\cdot} \theta_s \,\mathrm{d}W_s\right)_T = \exp\left(-\int_0^T \theta_s \,\mathrm{d}W_s - \tfrac{1}{2}\int_0^T \theta_s^2 \,\mathrm{d}s\right).$$

By the Itô isometry, (Z_t) is a non-negative local martingale with $\mathbb{E}[Z_0] = 1$. For it to be a true martingale (so that Z_T defines a valid probability measure), a sufficient condition is Novikov's:

Theorem: Novikov's Condition

If $\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \theta_s^2 \,\mathrm{d}s\right)\right] < \infty$, then $(Z_t)_{t\in[0,T]}$ is a uniformly integrable martingale.

9.3 Girsanov's Theorem

Theorem: Girsanov (one-dimensional)

Let $(\theta_t)_{t\in[0,T]}$ satisfy Novikov's condition and define $\mathrm{d}\mathbb{Q} = Z_T\,\mathrm{d}\mathbb{P}$. Then under $\mathbb{Q}$,

$$\tilde{W}_t := W_t + \int_0^t \theta_s\,\mathrm{d}s, \quad t \in [0,T],$$

is a standard Brownian motion.

Sketch. By Lévy's characterisation, it suffices to show $(\tilde{W}_t)$ is a continuous local $\mathbb{Q}$-martingale with $[\tilde{W}]_t = t$. The quadratic variation is immediate since $[\tilde{W}]_t = [W]_t = t$. For the martingale property, note that for $s < t$:

$$\mathbb{E}^{\mathbb{Q}}[\tilde{W}_t \mid \mathcal{F}_s] = \frac{1}{Z_s}\mathbb{E}^{\mathbb{P}}[Z_t\tilde{W}_t \mid \mathcal{F}_s],$$

and the Itô product rule $\mathrm{d}(Z_t\tilde{W}_t) = Z_t\,\mathrm{d}\tilde{W}_t + \tilde{W}_t\,\mathrm{d}Z_t + \mathrm{d}[Z,\tilde{W}]_t$ shows $Z_t\tilde{W}_t$ is a $\mathbb{P}$-local martingale. □

Application: Removal of drift

Consider $\mathrm{d}X_t = b(t,X_t)\,\mathrm{d}t + \mathrm{d}W_t$. Taking $\theta_t = b(t,X_t)$, the Girsanov measure $\mathbb{Q}$ turns $\tilde{W}_t = W_t + \int_0^t b(s,X_s)\,\mathrm{d}s$ into a Brownian motion, and under $\mathbb{Q}$ we have $\mathrm{d}X_t = \mathrm{d}\tilde{W}_t$, i.e. (X_t) is itself a $\mathbb{Q}$-Brownian motion. This is the key step in proving weak existence of SDEs via the Girsanov method.

9.4 Multidimensional Girsanov

Theorem: Girsanov (multidimensional)

Let $\mathbf{W} = (W^1,\dots,W^m)$ be an m-dimensional $\mathbb{P}$-Brownian motion, $\boldsymbol{\theta} = (\theta^1,\dots,\theta^m)$ a progressively measurable vector process satisfying $\mathbb{E}\Big[\exp\Big(\frac{1}{2}\int_0^T |\boldsymbol{\theta}_s|^2\,\mathrm{d}s\Big)\Big] < \infty$. Define $\mathrm{d}\mathbb{Q} = Z_T\,\mathrm{d}\mathbb{P}$ with

$$Z_T = \exp\left(-\sum_{j=1}^m \int_0^T \theta_s^j\,\mathrm{d}W_s^j - \frac{1}{2}\int_0^T |\boldsymbol{\theta}_s|^2\,\mathrm{d}s\right).$$

Then $\tilde{\mathbf{W}}_t = \mathbf{W}_t + \int_0^t \boldsymbol{\theta}_s\,\mathrm{d}s$ is a $\mathbb{Q}$-Brownian motion.

9.5 Relative Entropy and the Entropic Cost of Measure Change

The Girsanov framework has a natural information-theoretic interpretation via relative entropy (Kullback–Leibler divergence), which measures the cost of changing from one probability measure to another.

Definition: Relative Entropy

The **relative entropy** (or **KL divergence**) of $\mathbb{Q}$ with respect to $\mathbb{P}$ on $(, \mathcal{F}_T)$ is

$$D_{\mathrm{KL}}(\mathbb{Q}\|\mathbb{P}) = \mathbb{E}^{\mathbb{Q}}\left[\log \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right] = \mathbb{E}^{\mathbb{P}}\left[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \log \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right].$$

Proposition: Entropy of the Girsanov Change of Measure

Under the Girsanov change of measure $\mathrm{d}\mathbb{Q} = Z_T \,\mathrm{d}\mathbb{P}$ with $Z_T = \mathcal{E}(-\int_0^{\cdot} \boldsymbol{\theta}_s \cdot \mathrm{d}\mathbf{W}_s)_T$,

$$D_{\mathrm{KL}}(\mathbb{Q}\|\mathbb{P}) = \tfrac{1}{2}\mathbb{E}^{\mathbb{Q}}\left[\int_0^T |\boldsymbol{\theta}_s|^2 \,\mathrm{d}s\right].$$

Proof. Since $\log Z_T = -\int_0^T \boldsymbol{\theta}_s \cdot \mathrm{d}\mathbf{W}_s - \frac{1}{2}\int_0^T |\boldsymbol{\theta}_s|^2 \,\mathrm{d}s$ and $\mathrm{d}\mathbb{Q}/\,\mathrm{d}\mathbb{P} = Z_T^{-1}$ (the sign depends on the convention; using $\mathrm{d}\mathbb{Q} = Z_T\,\mathrm{d}\mathbb{P}$ means $\mathrm{d}\mathbb{Q}/\,\mathrm{d}\mathbb{P} = Z_T$), we compute

$$D_{\mathrm{KL}}(\mathbb{Q}\|\mathbb{P}) = \mathbb{E}^{\mathbb{Q}}[\log Z_T^{-1}] = -\mathbb{E}^{\mathbb{Q}}\left[-\int_0^T \theta_s \,\mathrm{d}\tilde{W}_s - \tfrac{1}{2}\int_0^T |\theta_s|^2 \,\mathrm{d}s + \int_0^T |\theta_s|^2 \,\mathrm{d}s\right]$$

where under $\mathbb{Q}$, $\tilde{W}$ is a Brownian motion and the stochastic integral has zero $\mathbb{Q}$-expectation, yielding $D_{\mathrm{KL}}(\mathbb{Q}\|\mathbb{P}) = \frac{1}{2}\mathbb{E}^{\mathbb{Q}}[\int_0^T |\theta_s|^2 \,\mathrm{d}s]$. □

Example: Optimal transport and the Schrödinger bridge

The problem of finding the measure $\mathbb{Q}^*$ on path space that minimises $D_{\mathrm{KL}}(\mathbb{Q}\|\mathbb{P}^W)$ subject to prescribed initial and terminal marginals μ_0 and μ_T is the **Schrödinger bridge** problem. By the entropic cost formula above, the optimal $\mathbb{Q}^*$ has drift θ^* minimising $\frac{1}{2}\mathbb{E}^{\mathbb{Q}}[\int_0^T |\theta_s|^2 \,\mathrm{d}s]$, which is an L^2 optimal transport problem with entropic regularisation. Schrödinger bridge methods have recently become the basis of a class of generative models in machine learning (diffusion models with stochastic interpolants).

Theorem: Donsker–Varadhan Variational Formula

For any measurable functional F of the Brownian path,

$$\log \mathbb{E}^{\mathbb{P}}[e^F] = \sup_{\mathbb{Q} \ll \mathbb{P}} \left\{\mathbb{E}^{\mathbb{Q}}[F] - D_{\mathrm{KL}}(\mathbb{Q}\|\mathbb{P})\right\}.$$

The supremum is attained by the tilted measure $\mathrm{d}\mathbb{Q}^* \propto e^F \,\mathrm{d}\mathbb{P}$.

This variational formula is the continuous-time analogue of the Donsker–Varadhan large-deviation result for Markov chains (Volume I, Chapter 20), and underlies the connection between stochastic optimal control and large deviation theory (the Freidlin–Wentzell theory, Chapter 21).

Chapter 10

The Martingale Representation Theorem

10.1 Statement and Proof

Theorem: Itô–Clark–Ocone Representation

Let $(B_t)_{t\in[0,T]}$ be a Brownian motion generating the filtration $(\mathcal{F}_t)$. For every $\mathcal{F}_T$-measurable random variable $\xi \in L^2(\mathbb{P})$, there exists a unique $H \in \mathcal{H}^2$ such that

$$\xi = \mathbb{E}[\xi] + \int_0^T H_s \, \mathrm{d}W_s.$$

Equivalently, every square-integrable martingale $M_t = \mathbb{E}[\xi \mid \mathcal{F}_t]$ is of the form

$$M_t = M_0 + \int_0^t H_s \, \mathrm{d}W_s.$$

Sketch. The set $\mathcal{M}$ of random variables representable as stochastic integrals (plus a constant) is a closed subspace of L^2 (by the Itô isometry). We show $\mathcal{M}^\perp = \{0\}$: if $\eta \perp \mathcal{M}$, then $\mathbb{E}[\eta e^{i\lambda B_T}] = 0$ for all $\lambda \in \mathbb{R}$ (since $e^{i\lambda B_T}$ lies in $\mathcal{M}$ by the stochastic exponential calculation). The set $\{e^{i\lambda B_T} : \lambda \in \mathbb{R}\}$ is total in L^2, forcing $\eta = 0$. $\square$

10.2 Consequences: Completeness of the Black–Scholes Market

The martingale representation theorem is the mathematical foundation of the replication argument in complete markets. In the Black–Scholes model, every $\mathcal{F}_T$-measurable payoff ξ can be replicated by a self-financing portfolio: the hedging strategy is $H_t = \partial_x u(t, S_t)$ where u solves the Black–Scholes PDE.

Example: European call option

A European call with strike K and maturity T has payoff $\xi = (S_T - K)^+$. Under the risk-neutral measure $\mathbb{Q}$ (obtained via Girsanov), $S_t = S_0 \exp(\sigma \tilde{W}_t - \frac{1}{2}\sigma^2 t)$. The arbitrage-free price at time t is

$$C(t, S_t) = e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}}[(S_T - K)^+ \mid \mathcal{F}_t] = S_t \Phi(d_+) - K e^{-r(T-t)} \Phi(d_-)$$

with $d_\pm = \frac{\ln(S_t/K) + (r \pm \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$, which is the **Black–Scholes formula**.

10.3 The Multidimensional Representation Theorem

Theorem: Representation under an m-Dimensional Brownian Filtration

Let $\mathbf{W} = (W^1, \dots, W^m)$ be an m-dimensional Brownian motion generating $(\mathcal{F}_t)_{t \in [0,T]}$. For every $\xi \in L^2(\mathcal{F}_T)$, there exists a unique predictable process $\mathbf{H} = (H^1, \dots, H^m) \in (\mathcal{H}^2)^m$ such that

$$\xi = \mathbb{E}[\xi] + \sum_{i=1}^{m} \int_0^T H_s^i \, \mathrm{d}W_s^i.$$

Remark: Market Completeness Requires Matching Dimensions

The multidimensional version of the theorem is precisely what underlies the classical statement that a Black–Scholes-type market with d traded risky assets driven by an m-dimensional Brownian motion is **complete** if and only if $d = m$ and the volatility matrix $\sigma(t, S_t) \in \mathbb{R}^{d \times m}$ is invertible (a.s., for a.e. t). If $d < m$ — fewer assets than independent sources of randomness — the market is **incomplete**: not every $\xi \in L^2(\mathcal{F}_T)$ can be replicated, and the representation $\xi = \mathbb{E}[\xi] + \int \mathbf{H} \cdot \mathrm{d}\mathbf{W}$ exists but $\mathbf{H}$ may not lie in the span of the traded assets' volatility vectors. This dichotomy is the starting point of the theory of incomplete markets and superhedging, beyond the scope of this volume.

10.4 The Clark–Ocone Formula

The representation theorem guarantees existence of the integrand H but the proof above gives no formula for it. Malliavin calculus (introduced briefly in Chapter 14) supplies an explicit expression.

Theorem: Clark–Ocone Formula

If $\xi \in \mathbf{D}^{1,2}$ (the Malliavin–Sobolev space, Chapter 14), then the integrand in the Itô–Clark–Ocone representation is given explicitly by

$$H_t = \mathbb{E}[D_t \xi \mid \mathcal{F}_t],$$

where $D_t \xi$ is the Malliavin derivative of ξ at time t. In particular,

$$\xi = \mathbb{E}[\xi] + \int_0^T \mathbb{E}[D_t \xi \mid \mathcal{F}_t] \, \mathrm{d}W_t.$$

Example: Hedging Ratio via Clark–Ocone

For $\xi = f(S_T)$ with $S_T = S_0 e^{(r-\sigma^2/2)T + \sigma W_T}$ (under $\mathbb{Q}$), the chain rule for Malliavin derivatives gives $D_t S_T = \sigma S_T$ (constant in t for $t \le T$), so $D_t \xi = f'(S_T)\sigma S_T$. The Clark–Ocone formula then recovers the familiar **delta-hedging** ratio

$$H_t = \sigma \, \mathbb{E}[f'(S_T) S_T \mid \mathcal{F}_t] = \sigma S_t \, \partial_x u(t, S_t),$$

matching, up to the factor σS_t, the PDE-based hedge $\partial_x u(t, S_t)$ stated in Section 10.2 — a striking instance of two independent derivations (probabilistic via Malliavin calculus, and analytic via the Feynman–Kac PDE) agreeing exactly.

Chapter 11

Stochastic Processes with Jumps: A Brief Overview

11.1 Poisson Processes and Compound Poisson Processes

Definition: Poisson Process

A **Poisson process** with rate $\lambda > 0$ is an adapted càdlàg process $(N_t)_{t\geq 0}$ with $N_0 = 0$, independent and stationary increments, and $N_t - N_s \sim \text{Poisson}(\lambda(t-s))$ for $s < t$. The compensated Poisson process $\tilde{N}_t := N_t - \lambda t$ is a martingale.

Proposition: Quadratic Variation of the Poisson Process

The compensated Poisson process $\tilde{N}_t = N_t - \lambda t$ has quadratic variation $[\tilde{N}]_t = N_t$ (a jump process: each jump of N contributes a unit jump to $[\tilde{N}]$), so $\mathbb{E}[[\tilde{N}]_t] = \lambda t$.

Definition: Compound Poisson Process

A **compound Poisson process** is $Y_t = \sum_{k=1}^{N_t} Z_k$ where (N_t) is a Poisson process with rate λ and $(Z_k)_{k\geq 1}$ are i.i.d. random variables with law ν (the jump size distribution), independent of N. Its characteristic function is $\mathbb{E}[e^{i\theta Y_t}] = \exp(t\lambda(\hat{\nu}(\theta) - 1))$ where $\hat{\nu}(\theta) = \int e^{i\theta z}\nu(\,\mathrm{d}z)$.

Example: Jump-diffusion for asset prices

The Merton (1976) model combines geometric Brownian motion with a compound Poisson jump component:

$$S_t = S_0 \exp\left(\left(\mu - \tfrac{\sigma^2}{2}\right)t + \sigma W_t + \sum_{k=1}^{N_t} \log(1 + Z_k)\right),$$

where $Z_k > -1$ are i.i.d. log-normal jump sizes. This models sudden large moves in asset prices (e.g., crashes), capturing fat tails in the return distribution absent from the Black–Scholes model. The risk-neutral option pricing for-

mula for Merton's model is an infinite series of Black–Scholes prices weighted by Poisson probabilities.

Definition: Lévy Process

A **Lévy process** (X_t) is an adapted càdlàg process with $X_0 = 0$, independent and stationary increments, and stochastic continuity. By the Lévy–Khintchine theorem, the characteristic function is

$$\mathbb{E}[e^{i\theta X_t}] = e^{t\psi(\theta)}, \quad \psi(\theta) = i\mu\theta - \tfrac{\sigma^2\theta^2}{2} + \int_{\mathbb{R}\setminus\{0\}} \left(e^{i\theta x} - 1 - i\theta x \mathbf{1}_{\{|x|\le 1\}}\right) \nu(\mathrm{d}x),$$

where $\mu \in \mathbb{R}$, $\sigma \ge 0$, and ν is the **Lévy measure** satisfying $\int (1 \wedge x^2)\nu(\mathrm{d}x) < \infty$.

The Lévy–Itô decomposition writes every Lévy process as the sum of a Brownian motion with drift, a compound Poisson process of large jumps ($|x| > 1$), and a square-integrable martingale of small jumps ($|x| \le 1$). This decomposition is canonical and the three components are independent.

Theorem: Lévy–Itô Decomposition

Every Lévy process (X_t) with triplet (μ, σ^2, ν) decomposes as

$$X_t = \mu t + \sigma W_t + \int_{|x|>1} x \tilde{N}(\mathrm{d}s, \mathrm{d}x) + \int_{|x|\le 1} x (N - \hat{N})(\mathrm{d}s, \mathrm{d}x),$$

where W is a standard Brownian motion, $N(\mathrm{d}s, \mathrm{d}x)$ is a Poisson random measure with intensity $\mathrm{d}s\,\nu(\mathrm{d}x)$, $\hat{N}(\mathrm{d}s, \mathrm{d}x) = \mathrm{d}s\,\nu(\mathrm{d}x)$ is the compensator, and all three components are independent.

Remark: Connection to Volume I

The Lévy–Khintchine formula is the continuous-time analogue of the Donsker–Varadhan large-deviation machinery of Volume I: the log-moment generating function $\psi(\theta)$ of a Lévy process is exactly the "Donsker–Varadhan functional" for the i.i.d. increments, and the Cramér transform $\psi_\lambda(\theta) = \psi(\theta + \lambda) - \psi(\lambda)$ corresponds to exponential tilting of the jump measure, paralleling the tilted transition matrix P_λ of Volume I Chapter 20.

11.2 The Poisson Stochastic Integral and Compensated Integral

For a progressively measurable function $f(t, x)$ and a Poisson random measure N with intensity $\hat{N}(\mathrm{d}t, \mathrm{d}x) = \mathrm{d}t\,\nu(\mathrm{d}x)$, the **stochastic integral with respect to the compen-**

sated measure is

$$\int_0^t \int_{\mathbb{R}\setminus\{0\}} f(s,x)\,\tilde{N}(\mathrm{d}s,\mathrm{d}x) := \int_0^t \int f(s,x)\,N(\mathrm{d}s,\mathrm{d}x) - \int_0^t \int f(s,x)\,\nu(\mathrm{d}x)\,\mathrm{d}s,$$

provided $\int_0^t \int f(s,x)^2 \nu(\mathrm{d}x)\,\mathrm{d}s < \infty$ a.s. This integral is a square-integrable martingale with quadratic variation $\int_0^t \int f(s,x)^2\nu(\mathrm{d}x)\,\mathrm{d}s$.

Theorem: Itô's Formula for Jump-Diffusions

Let $X_t = X_0 + \int_0^t b_s\,\mathrm{d}s + \int_0^t \sigma_s\,\mathrm{d}W_s + \int_0^t \int_{\mathbb{R}\setminus\{0\}} x\,\tilde{N}(\mathrm{d}s,\mathrm{d}x)$ and $f \in C^2$. Then

$$f(X_t) = f(X_0) + \int_0^t f'(X_{s-})\,\mathrm{d}X_s^c + \tfrac{1}{2}\int_0^t f''(X_{s-})\sigma_s^2\,\mathrm{d}s + \sum_{0<s\le t}\big[f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\big],$$

where X^c is the continuous part and $\Delta X_s = X_s - X_{s-}$ are the jumps.

A **Lévy process** (X_t) is an adapted càdlàg process with $X_0 = 0$, independent and stationary increments, and stochastic continuity. By the Lévy–Khintchine theorem, the characteristic function is

$$\mathbb{E}[e^{i\theta X_t}] = e^{t\psi(\theta)}, \quad \psi(\theta) = i\mu\theta - \tfrac{\sigma^2\theta^2}{2} + \int_{\mathbb{R}\setminus\{0\}} \big(e^{i\theta x} - 1 - i\theta x \mathbf{1}_{\{|x|\le 1\}}\big)\,\nu(\mathrm{d}x),$$

where $\mu \in \mathbb{R}$, $\sigma \ge 0$, and ν is the **Lévy measure** satisfying $\int (1 \wedge x^2)\nu(\mathrm{d}x) < \infty$.

The Lévy–Itô decomposition writes every Lévy process as the sum of a Brownian motion with drift, a compound Poisson process, and a martingale of small jumps. Jump-diffusion SDEs driven by Lévy processes are treated briefly in Chapter 22.

Part III

Stochastic Differential Equations

Chapter 12

Existence and Uniqueness of SDEs

12.1 Formulation

Definition: Stochastic Differential Equation

A **stochastic differential equation** (SDE) on $\mathbb{R}^d$ driven by an m-dimensional Brownian motion $\mathbf{W}$ is

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}\mathbf{W}_t, \quad X_0 = x_0 \in \mathbb{R}^d, \tag{12.1}$$

where $b\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ is the **drift**, $\sigma\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}^{d\times m}$ is the **diffusion coefficient**, and the equation is understood in integral form:

$$X_t = x_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}\mathbf{W}_s.$$

A **strong solution** is a process (X_t) that is adapted to the filtration generated by $\mathbf{W}$ and satisfies the integral equation a.s. A **weak solution** relaxes the requirement that X be measurable with respect to the given Brownian motion; it allows constructing a new probability space where the equation is satisfied.

12.2 Lipschitz Theory: Itô's Theorem

Theorem: Itô's Existence and Uniqueness Theorem

Suppose b and σ are measurable and satisfy the **Lipschitz** and **linear growth** conditions: there exists $C > 0$ such that for all $t \in [0,T]$, $x, y \in \mathbb{R}^d$,

$$|b(t,x) - b(t,y)| + \|\sigma(t,x) - \sigma(t,y)\| \le C\,|x-y|\,, \tag{12.2}$$

$$|b(t,x)|^2 + \|\sigma(t,x)\|^2 \le C^2(1+|x|^2). \tag{12.3}$$

Then (12.1) has a unique strong solution $(X_t)_{t\in[0,T]}$ (unique up to indistinguishability). Moreover, $\sup_{t\in[0,T]} \mathbb{E}[|X_t|^2] < \infty$.

Sketch: Picard iteration. Define $X_t^0 \equiv x_0$ and

$$X_t^{n+1} = x_0 + \int_0^t b(s, X_s^n)\,\mathrm{d}s + \int_0^t \sigma(s, X_s^n)\,\mathrm{d}W_s.$$

Using the Lipschitz condition, the Itô isometry, and Gronwall's inequality, one shows

$$\mathbb{E}\left[\sup_{s\le t} |X_s^{n+1} - X_s^n|^2\right] \le K \int_0^t \mathbb{E}\left[\sup_{r\le s} |X_r^n - X_r^{n-1}|^2\right] \mathrm{d}s,$$

so the sequence converges geometrically in L^2 to a limit X that satisfies (12.1). Uniqueness follows by applying Gronwall to the difference of two solutions. □

Remark

The Lipschitz condition (12.2) is sufficient but not necessary for existence and uniqueness. Important cases beyond Lipschitz include:

(i) **One-dimensional, monotone drift**: b non-increasing (or satisfying $xb(x) \le C(1+x^2)$) with σ Lipschitz;

(ii) **Hörmander condition**: σ Lipschitz, b merely measurable – existence of a weak solution via Girsanov;

(iii) **Yamada–Watanabe**: one-dimensional SDEs with Hölder drift $|b(x) - b(y)| \le C|x-y|^\alpha$, $\alpha \ge 1/2$.

12.3 Moment Estimates

Theorem: p-th Moment Bound

Under the linear growth condition (12.3) and for $p \ge 2$, the strong solution satisfies

$$\mathbb{E}\left[\sup_{0\le t\le T} |X_t|^p\right] \le C_p(1+|x_0|^p)e^{C_p T}$$

for a constant C_p depending only on p, T, and the constant C in (12.3).

Proof. Apply Itô's formula to $|X_t|^p$, use the linear growth condition to bound the stochastic integral, apply Doob's maximal inequality, and conclude by Gronwall. □

12.4 Continuity in Initial Conditions

Proposition

Under the Lipschitz condition (12.2), if X^x and X^y denote the strong solutions starting from x and y respectively, then

$$\mathbb{E}\left[\sup_{0\le t\le T}|X_t^x - X_t^y|^2\right] \le C_T\,|x-y|^2\,.$$

In particular, $x \mapsto X_t^x$ is continuous (in L^2 and a.s.) for each t.

This flow property is essential for treating the SDE solution as a stochastic flow of diffeomorphisms, discussed in Chapter 14.

Chapter 13

The Markov Property of SDE Solutions

13.1 The Strong Markov Property

Theorem: SDE solutions are strong Markov

Under the Lipschitz condition, the solution $(X_t^{s,x})_{t\geq s}$ to (12.1) with $X_s = x$ defines a Markov family with transition semigroup

$$P_{s,t}f(x) := \mathbb{E}[f(X_t^{s,x})].$$

If coefficients are time-homogeneous ($b(t,x) = b(x)$, $\sigma(t,x) = \sigma(x)$), then (X_t) is a homogeneous Markov process with semigroup (P_t) satisfying $P_{t+s} = P_t P_s$.

Theorem

Under the Lipschitz condition, (X_t) has the **strong Markov property**: for any stopping time τ, the conditional distribution of $(X_{\tau+t})_{t\geq 0}$ given $\mathcal{F}_\tau$ equals the distribution of $(X_t^{X_\tau})_{t\geq 0}$.

13.2 The Generator of a Diffusion

Definition: Infinitesimal Generator

The **generator** $\mathcal{L}$ of the homogeneous SDE $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$ acts on $f \in C^2(\mathbb{R}^d)$ by

$$\mathcal{L}f(x) = \sum_i b^i(x)\partial_{x_i} f(x) + \tfrac{1}{2}\sum_{i,k} a^{ik}(x)\partial^2_{x_i x_k} f(x),$$

where $a = \sigma\sigma^\top$ is the **diffusion matrix**.

Dynkin's Formula

For $f \in C_b^2(\mathbb{R}^d)$ and any stopping time τ with $\mathbb{E}[\tau] < \infty$,

$$\mathbb{E}^x[f(X_\tau)] = f(x) + \mathbb{E}^x\left[\int_0^\tau \mathcal{L}f(X_s)\,\mathrm{d}s\right].$$

Example: Brownian motion generator

For $\mathrm{d}X_t = \mathrm{d}W_t$ (i.e. $b \equiv 0$, $\sigma = I$), the generator is $\mathcal{L} = \tfrac{1}{2}\Delta$, the half-Laplacian. Dynkin's formula then reads $\mathbb{E}^x[f(B_\tau)] = f(x) + \tfrac{1}{2}\mathbb{E}^x[\int_0^\tau \Delta f(B_s)\,\mathrm{d}s]$, recovering the mean value property of harmonic functions when $\mathcal{L}f = 0$.

13.3 The Kolmogorov Equations and the Semigroup

The generator $\mathcal{L}$ encodes the local (infinitesimal) behaviour of the diffusion; the semigroup (P_t) encodes its global evolution. The two are linked by a pair of PDEs that play a role analogous to the forward and backward equations for Markov chains in Volume I.

Theorem: Kolmogorov Backward Equation

Let $u(t,x) := P_{T-t}f(x) = \mathbb{E}^x[f(X_{T-t})]$ for fixed T and $f \in C_b^2(\mathbb{R}^d)$. Then u solves the **backward Kolmogorov equation**

$$\partial_t u(t,x) + \mathcal{L}u(t,x) = 0, \qquad u(T,x) = f(x),$$

where $\mathcal{L}$ acts on the spatial variable x.

Sketch. By the Markov property and the tower rule, $u(t, X_t) = \mathbb{E}[f(X_T) \mid \mathcal{F}_t]$ is a martingale. Applying Itô's formula to $u(t, X_t)$ and using that the drift term of a martingale must vanish gives exactly $\partial_t u + \mathcal{L}u = 0$; this is precisely the route used to prove the Feynman–Kac formula in Chapter 17, of which this is the special (undiscounted, unkilled) case. □

Remark: Forward Equation Preview

The companion **forward** (Fokker–Planck) equation describes the evolution of the *density* $\rho(t,x)$ of X_t rather than the expectation of a fixed observable, and involves the formal adjoint operator $\mathcal{L}^*$ rather than $\mathcal{L}$ itself. The forward/backward duality $\langle \mathcal{L}f, \rho\rangle = \langle f, \mathcal{L}^*\rho\rangle$ is developed fully in Chapter 17; the present chapter's generator $\mathcal{L}$ is exactly the operator appearing on the backward side of that duality.

Semigroup Properties

The family $(P_t)_{t\geq 0}$ defined by $P_t f(x) = \mathbb{E}^x[f(X_t)]$ satisfies:

(i) $P_0 = \mathrm{Id}$ and $P_{t+s} = P_t P_s$ (the semigroup property, a restatement of the Markov property via the Chapman–Kolmogorov relation);

(ii) $\|P_t f\|_\infty \leq \|f\|_\infty$ (contraction on C_b);

(iii) $\partial_t P_t f\big|_{t=0} = \mathcal{L} f$ for f in the domain of $\mathcal{L}$ (this is the defining property of the generator, in the Hille–Yosida sense);

(iv) if b, σ are bounded and Lipschitz, (P_t) is a strongly continuous (Feller) semigroup on $C_0(\mathbb{R}^d)$.

Chapter 14

Stability and Lyapunov Methods for SDEs

14.1 Notions of Stability

For the time-homogeneous SDE $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$, the equilibrium $x^* = 0$ (assuming $b(0) = \sigma(0) = 0$) is:

(a) **Stable in probability** if for every $\varepsilon, \eta > 0$ there exists $\delta > 0$ such that $|x| < \delta \Rightarrow \mathbb{P}(\sup_{t \geq 0} |X_t^x| > \varepsilon) < \eta$;

(b) **Asymptotically stable** if stable and $X_t^x \to 0$ in probability as $t \to \infty$ for all sufficiently small x;

(c) **Exponentially stable** if $\mathbb{E}[|X_t^x|^p] \leq C\,|x|^p\, e^{-\lambda t}$ for some $\lambda, C > 0$.

14.2 Stochastic Lyapunov Functions

Theorem: Khasminskii Stability Criterion

Suppose there exists $V \in C^2(\mathbb{R}^d)$ with $V(0) = 0$, $V(x) > 0$ for $x \neq 0$, $V(x) \to \infty$ as $|x| \to \infty$, and

$$\mathcal{L}V(x) \leq -cV(x), \quad c > 0,$$

for all x in a neighbourhood of 0. Then the equilibrium $x^* = 0$ is exponentially stable.

Example: Ornstein–Uhlenbeck

For $\mathrm{d}X_t = -\theta X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$, take $V(x) = x^2$. Then

$$\mathcal{L}V(x) = -2\theta x^2 + \sigma^2 = -2\theta V(x) + \sigma^2.$$

This is not exactly $\leq -cV(x)$, but $\mathcal{L}V \leq -2\theta V + \sigma^2 \leq -\theta V$ for $V \geq \sigma^2/\theta$. The drift dominates outside a compact set; the equilibrium is not zero in the strict

Lyapunov sense, but the process is ergodic with a unique stationary distribution $\mathcal{N}(0, \sigma^2/(2\theta))$.

14.3 Explosion and Non-explosion

Khasminskii Non-explosion Criterion

The solution of $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$ is non-explosive (defined for all $t \geq 0$) if there exists $V \in C^2(\mathbb{R}^d)$ with $V(x) \to \infty$ and $\mathcal{L}V(x) \leq CV(x)$ for some constant C.

14.4 The Lyapunov Exponent and Almost-Sure Stability

The exponential stability criterion above bounds $\mathbb{E}[|X_t|^p]$, a statement about *moments*. A finer and, in some respects, more natural notion concerns the a.s. exponential growth rate of individual paths.

Definition: Lyapunov Exponent

For the SDE $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$ with $X_0 = x \neq 0$, the (top) **Lyapunov exponent** is

$$\lambda := \limsup_{t \to \infty} \frac{1}{t} \ln |X_t| \,.$$

The equilibrium is **almost-surely exponentially stable** if $\lambda < 0$ a.s. for every $x \neq 0$.

Example: Lyapunov Exponent of the OU Process

For the Ornstein–Uhlenbeck process $\mathrm{d}X_t = -\theta X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$, the explicit solution $X_t = xe^{-\theta t} + \sigma \int_0^t e^{-\theta(t-s)}\,\mathrm{d}W_s$ has $\frac{1}{t}\ln|X_t| \to -\theta$ a.s. as $t \to \infty$ (the stochastic integral term grows only like $\sqrt{t}\log\log t$ by the law of the iterated logarithm, Chapter 3, which is negligible against the linear decay $-\theta t$). Hence $\lambda = -\theta < 0$: a.s. exponential stability holds with rate exactly θ, matching the L^p rate found via the Lyapunov function method above.

Remark: A.s. versus Moment Stability Can Differ

For *multiplicative* (state-dependent) noise, a.s. and moment stability can genuinely disagree. Consider the scalar SDE $\mathrm{d}X_t = \alpha X_t\,\mathrm{d}t + \beta X_t\,\mathrm{d}W_t$ (geometric Brownian motion with drift α). By Itô's formula applied to $\ln X_t$ (Chapter 9, Exercise 10), $\ln X_t = \ln x + (\alpha - \beta^2/2)t + \beta W_t$, so the Lyapunov exponent is exactly $\lambda = \alpha - \beta^2/2$ a.s. Yet $\mathbb{E}[X_t] = xe^{\alpha t}$, so the *first moment* grows whenever $\alpha > 0$, even though the path itself decays a.s. to 0 whenever $\alpha < \beta^2/2$.

This is the same dichotomy noted for GBM in Exercise 10: the mean is dominated by rare, exponentially large excursions, while almost every path follows the smaller a.s. rate $\alpha - \beta^2/2$. The moral is that Lyapunov-function criteria of the type in Theorem 16.4, which control moments, do not automatically imply, nor are implied by, a.s. stability; the two notions require separate verification whenever p-th moment growth rates depend on p.

Theorem: A.s. Stability via the Logarithmic Transformation

Suppose σ is bounded away from 0 and that there is a function $V \in C^2(\mathbb{R}^d \setminus \{0\})$, $V(x) = \ln|x| + W(x)$ with W bounded, such that $\mathcal{L}V(x) \le -c$ for some $c > 0$ and all $x \neq 0$ in a punctured neighbourhood of the origin. Then $\lambda \le -c$ a.s.: the equilibrium is almost-surely exponentially stable.

Sketch. Apply Itô's formula to $V(X_t)$ to obtain $V(X_t) = V(x) + \int_0^t \mathcal{L}V(X_s)\,\mathrm{d}s + M_t$ for a local martingale M_t with bracket $[M]_t = \int_0^t |\nabla V \sigma|^2(X_s)\,\mathrm{d}s$, bounded linearly in t under the stated hypotheses. The strong law of large numbers for martingales (an extension of Theorem 2.16 to local martingales with $[M]_t/t \to$ a finite limit) gives $M_t/t \to 0$ a.s., so $V(X_t)/t \to -c$ or smaller, and since W is bounded, $\ln|X_t|/t$ inherits the same limsup. $\square$

Chapter 15

Numerical Methods for SDEs

15.1 The Euler–Maruyama Scheme

Definition: Euler–Maruyama

For the SDE $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$ with $X_0 = x_0$ on $[0,T]$, the **Euler–Maruyama** (EM) scheme with stepsize $h = T/N$ is

$$X^h_{(k+1)h} = X^h_{kh} + b(X^h_{kh})h + \sigma(X^h_{kh})(W_{(k+1)h} - W_{kh}), \quad k = 0,1,\dots,N-1.$$

Theorem: Strong Convergence of EM

Under global Lipschitz and linear growth conditions,

$$\mathbb{E}\left[\sup_{0\le t\le T} |X_t - X^h_t|^2\right]^{1/2} \le C h^{1/2},$$

i.e. the EM scheme has **strong order** $1/2$.

15.2 The Milstein Scheme

By including the next term in the stochastic Taylor expansion, one obtains a higher-order method:

Definition: Milstein Scheme

The **Milstein scheme** is

$$X^h_{(k+1)h} = X^h_{kh} + b(X^h_{kh})h + \sigma(X^h_{kh})\Delta W_k + \sigma'(X^h_{kh})\sigma(X^h_{kh})\tfrac{1}{2}[(\Delta W_k)^2 - h],$$

where $\Delta W_k = W_{(k+1)h} - W_{kh}$ and $\sigma' = \partial_x \sigma$ (in one dimension).

Theorem: Strong Convergence of Milstein

Under appropriate smoothness conditions, the Milstein scheme has **strong order 1**:

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|X_t - X_t^h\right|^2\right]^{1/2} \le Ch.$$

15.3 The Stochastic Taylor Expansion

The hierarchy of numerical schemes for SDEs is most systematically understood through the **stochastic Taylor expansion** (also called the Itô–Wagner–Platen expansion).

Definition: Multiple Stochastic Integrals

For a multi-index $\alpha = (\alpha_1, \dots, \alpha_k)$ with $\alpha_i \in \{0, 1\}$ (where 0 denotes integration in $\mathrm{d}s$ and 1 denotes integration in $\mathrm{d}W_s$), define the **iterated stochastic integral**

$$I_\alpha[f]_{s,t} = \int_s^t \int_s^{r_k} \cdots \int_s^{r_2} f(r_1)\, \mathrm{d}r_1^{(\alpha_1)} \cdots \mathrm{d}r_k^{(\alpha_k)},$$

where $\mathrm{d}r^{(0)} = \mathrm{d}r$ and $\mathrm{d}r^{(1)} = \mathrm{d}W_r$.

Theorem: Stochastic Taylor Expansion

For the SDE $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$ with $b, \sigma \in C_b^\infty$, the solution satisfies

$$\begin{aligned} X_t = X_s &+ b(X_s)(t-s) + \sigma(X_s)(W_t - W_s) \\ &+ \sigma'(X_s)\sigma(X_s)\, I_{(1,1)}[1]_{s,t} + \big(b'(X_s)\sigma(X_s) + \tfrac{1}{2}\sigma''(X_s)\sigma(X_s)^2\big) I_{(1,0)}[1]_{s,t} + R_{s,t}, \end{aligned}$$

where $I_{(1,1)}[1]_{s,t} = \int_s^t (W_r - W_s)\,\mathrm{d}W_r = \tfrac{1}{2}[(W_t - W_s)^2 - (t-s)]$ and the remainder $R_{s,t} = O((t-s)^{3/2})$ in L^2.

Retaining only the terms through order $h^{1/2}$ (i.e., the first three) gives the EM scheme (strong order $1/2$). Retaining terms through order h gives the Milstein scheme (strong order 1). Going to order $h^{3/2}$ yields the **order-1.5 Taylor scheme**, which requires simulating the joint law of ΔW_k and $I_{(1,0)}[1]$, a correlated Gaussian pair.

Remark: Commutative and Non-commutative Cases

In the multidimensional case $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sum_j \sigma^j(X_t)\,\mathrm{d}W_t^j$, the Milstein correction involves terms $\partial_{x_l}\sigma^j \sigma^{j'} I_{(j,j')}[1]$ for $j \neq j'$, which require simulating Lévy areas (double integrals $\int_s^t W_{r,1}\,\mathrm{d}W_{r,2}$). When σ^j and $\sigma^{j'}$ commute in the sense that $\partial_{x_l}\sigma^{j,i}\sigma^{l,j'} = \partial_{x_l}\sigma^{j',i}\sigma^{l,j}$ for all i, j, j', the Lévy area terms cancel and the multidimensional Milstein scheme is easily implementable. This **commutativity condition** holds, for example, when $\sigma(x) = \text{const.}$ or when $d = m = 1$.

15.4 Runge–Kutta Methods and Weak Convergence

For the purpose of approximating $\mathbb{E}[f(X_T)]$ (rather than individual trajectories), **weak convergence** suffices: $\left|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^h)]\right| \le Ch^p$ for some order p.

Theorem

The Euler–Maruyama scheme has **weak order 1**. There exist explicit Runge–Kutta-type schemes achieving weak order 2, which are often preferred in practice (e.g. for Monte Carlo simulation of option prices) as they attain higher accuracy without requiring derivatives of σ.

Example: Weak order-2 Runge–Kutta (Platen)

A simple weak order-2 scheme for the scalar SDE $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$ is:

$$\begin{aligned} \hat{X} &= X_{kh} + b(X_{kh})h + \sigma(X_{kh})\sqrt{h}, \\ X_{(k+1)h} &= X_{kh} + \tfrac{1}{2}[b(\hat{X}) + b(X_{kh})]h + \tfrac{1}{2}[\sigma(\hat{X}) + \sigma(X_{kh})]\Delta W_k \\ &\quad + \tfrac{1}{2\sqrt{h}}[\sigma(\hat{X}) - \sigma(X_{kh})][(\Delta W_k)^2 - h]. \end{aligned}$$

This scheme requires only one standard Gaussian variate per step (same as EM) but achieves weak order 2 under C_b^4 smoothness of b, σ, f.

Algorithm 1: Euler–Maruyama simulation of an SDE on $[0, T]$

Input: Drift b, diffusion σ, initial condition x_0, number of steps N, T
Output: Approximate path $\{X_{kh}^h\}_{k=0}^N$

```
h ← T/N;
X_0 ← x_0;
for k = 0, 1, ..., N − 1 do
    Sample ΔW ~ N(0, h);
    X_{(k+1)h} ← X_{kh} + b(X_{kh}) · h + σ(X_{kh}) · ΔW;
return {X_{kh}};
```

15.5 Stochastic Runge–Kutta and Implicit Methods

For stiff SDEs (large negative eigenvalues of $\partial_x b$), explicit methods require very small stepsizes. **Implicit** methods solve a nonlinear equation at each step but allow much larger h:

$$X_{(k+1)h}^h = X_{kh}^h + b(X_{(k+1)h}^h)h + \sigma(X_{kh}^h)\Delta W_k \quad \text{(semi-implicit EM)}.$$

The semi-implicit EM has the same strong order $1/2$ as the explicit EM but better stability properties. For the OU process with $b(x) = -\theta x$ and large θ, the semi-implicit scheme is unconditionally stable in the mean-square sense.

Proposition: Mean-Square Stability of Semi-Implicit EM

For the linear SDE $\mathrm{d}X_t = -\theta X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$ with $\theta > 0$, the semi-implicit Euler scheme $X_{k+1} = X_k(1+\theta h)^{-1} + \sigma\Delta W_k(1+\theta h)^{-1}$ satisfies $\mathbb{E}[X_k^2] \to \sigma^2/(2\theta)$ as $k \to \infty$ for any $h > 0$, reproducing the correct stationary variance. In contrast, the explicit EM requires $h < 1/\theta$ for stability.

Chapter 16

The Stochastic Flow and Malliavin Calculus (Introduction)

16.1 The Stochastic Flow of Diffeomorphisms

Under Lipschitz conditions, the solution map $x \mapsto X_t^x$ defines, for each t, an a.s. continuous map. Under additional smoothness of b and σ (e.g. C_b^k), Kunita's theorem asserts that this map is a.s. a C^k diffeomorphism:

Kunita's Flow Theorem

If $b, \sigma \in C_b^k$ for $k \geq 1$, the map $\phi_t \colon x \mapsto X_t^x$ is a.s. a C^k diffeomorphism of $\mathbb{R}^d$ for every t. The derivative flow $D\phi_t = (\partial_{x_j} X_t^{x,i})$ satisfies the linearised SDE

$$\mathrm{d}D\phi_t = \partial_x b(X_t^x) D\phi_t \, \mathrm{d}t + \sum_j \partial_x \sigma^j(X_t^x) D\phi_t \, \mathrm{d}W_t^j, \quad D\phi_0 = I.$$

16.2 Introduction to Malliavin Calculus

Malliavin calculus extends the classical differential calculus to functionals of Brownian motion, providing a probabilistic proof of Hörmander's theorem on hypoelliptic operators and an explicit representation of hedging strategies in finance.

Definition: Malliavin Derivative

For a smooth Brownian functional $F = f(B_{t_1}, \ldots, B_{t_n})$ with $f \in C_b^\infty(\mathbb{R}^n)$, the **Malliavin derivative** is the $L^2([0,T])$-valued random variable

$$D_s F = \sum_{i=1}^{n} \partial_{x_i} f(B_{t_1}, \ldots, B_{t_n}) \mathbf{1}_{[0,t_i]}(s).$$

The domain $\mathbf{D}^{1,2}$ of D is the closure of smooth functionals in L^2 under the Sobolev norm $\|F\|_{1,2}^2 = \mathbb{E}[F^2] + \mathbb{E}[\int_0^T (D_s F)^2 \, \mathrm{d}s]$. The adjoint $D^* = \delta$ is the Skorokhod integral, which extends the Itô integral to non-adapted integrands.

Clark–Ocone Formula

For $F \in \mathbf{D}^{1,2}$,

$$F = \mathbb{E}[F] + \int_0^T \mathbb{E}[D_s F \mid \mathcal{F}_s]\, \mathrm{d}W_s.$$

The Clark–Ocone formula gives an explicit representation of the hedging strategy: for a payoff $\xi = F(S_T)$ in a diffusion market, the replicating strategy at time s is $\mathbb{E}[D_s F \mid \mathcal{F}_s]/S_s\sigma$.

16.3 The Chain Rule and the Skorokhod Integral

Chain Rule for the Malliavin Derivative

For $\phi \in C_b^1(\mathbb{R})$ and $F \in \mathbf{D}^{1,2}$, $\phi(F) \in \mathbf{D}^{1,2}$ and

$$D_s \phi(F) = \phi'(F)\, D_s F.$$

More generally, for $\mathbf{F} = (F^1, \dots, F^n) \in (\mathbf{D}^{1,2})^n$ and $\phi \in C_b^1(\mathbb{R}^n)$, $D_s\phi(\mathbf{F}) = \sum_i \partial_i \phi(\mathbf{F}) D_s F^i$.

Definition: Skorokhod Integral

The **Skorokhod integral** $\delta(u)$ of a (possibly non-adapted) process $u \in L^2([0,T]\times\Omega)$ is defined as the adjoint of D: it is the unique element of $L^2(\Omega)$ (when it exists) satisfying the duality relation

$$\mathbb{E}[F\,\delta(u)] = \mathbb{E}\left[\int_0^T D_s F \cdot u_s \, \mathrm{d}s\right] \qquad \text{for all } F \in \mathbf{D}^{1,2}.$$

Theorem: Skorokhod Extends the Itô Integral

If u is adapted and $u \in \mathcal{H}^2$, then $\delta(u) = \int_0^T u_s \, \mathrm{d}W_s$, the ordinary Itô integral. In particular, the Skorokhod integral coincides with the Itô integral on adapted integrands but is defined more generally for anticipating (non-adapted) integrands as well.

Integration by Parts Formula

For $F \in \mathbf{D}^{1,2}$ and u in the domain of δ,

$$\delta(Fu) = F\,\delta(u) - \int_0^T D_s F\, u_s \, \mathrm{d}s,$$

provided Fu is itself in the domain of δ. This Leibniz-type rule is the source of

the **Malliavin integration by parts formula**: for $F, G \in \mathbf{D}^{1,2}$ and $u \in \mathcal{H}^2$,

$$\mathbb{E}[G\,\delta(Fu)] = \mathbb{E}[FG\,\delta(u)] - \mathbb{E}\left[G\int_0^T D_s F\,u_s\,\mathrm{d}s\right].$$

16.4 The Malliavin Covariance Matrix and Hörmander's Theorem

The deepest application of Malliavin calculus in this volume is a probabilistic criterion for the existence of a smooth density for X_t, even when the diffusion matrix $a = \sigma\sigma^\top$ degenerates (so that classical elliptic PDE theory does not directly apply).

Definition: Malliavin Covariance Matrix

For $X_t \in \mathbb{R}^d$ solving the SDE $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$ with smooth coefficients, the **Malliavin covariance matrix** is

$$\gamma_{X_t} := \int_0^t D_s X_t\,(D_s X_t)^\top\,\mathrm{d}s \in \mathbb{R}^{d\times d}.$$

Theorem: Existence of a Smooth Density (Malliavin's Criterion)

If $X_t \in (\mathbf{D}^{1,2})^d$ and γ_{X_t} is invertible a.s. with $\det(\gamma_{X_t})^{-1} \in \bigcap_{p\geq 1} L^p$, then the law of X_t has a C^∞ density with respect to Lebesgue measure on $\mathbb{R}^d$.

Remark: Hörmander's Condition

For SDEs of the form $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sum_{j=1}^m \sigma_j(X_t)\,\mathrm{d}W_t^j$ with smooth vector fields $b, \sigma_1, \ldots, \sigma_m$, Hörmander's bracket condition requires that the vector fields $\sigma_1, \ldots, \sigma_m$ together with their iterated Lie brackets $[\sigma_i, \sigma_j]$, $[\sigma_i, [\sigma_j, \sigma_k]]$, ...span $\mathbb{R}^d$ at every point. This condition — weaker than ellipticity of $a = \sigma\sigma^\top$, since it allows σ to have a non-trivial kernel as long as brackets fill in the missing directions — implies invertibility of γ_{X_t} via Malliavin's criterion, and hence existence of a smooth density even for genuinely degenerate diffusions. This is Malliavin's celebrated probabilistic proof of Hörmander's hypoellipticity theorem, originally established by purely analytic (pseudo-differential) methods.

Part IV

Fokker–Planck and Kolmogorov Equations

Chapter 17

The Fokker–Planck Equation

17.1 Derivation from the SDE

Consider the time-homogeneous SDE $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$ with diffusion matrix $a = \sigma\sigma^{\top}$. Let $\rho(t,x)$ denote the density of X_t (assuming it exists). We derive the equation governing the evolution of ρ.

For any test function $\phi \in C_c^\infty(\mathbb{R}^d)$, Dynkin's formula gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\phi(X_t)] = \mathbb{E}[\mathcal{L}\phi(X_t)] = \int_{\mathbb{R}^d} \mathcal{L}\phi(x)\rho(t,x)\,\mathrm{d}x.$$

Integrating $\mathcal{L}\phi$ by parts (transferring the differential operators from ϕ to ρ):

$$\int \mathcal{L}\phi\rho\,\mathrm{d}x = \int \phi\mathcal{L}^*\rho\,\mathrm{d}x,$$

where $\mathcal{L}^*$ is the L^2-adjoint of the generator. Since $\frac{\mathrm{d}}{\mathrm{d}t}\int \phi\rho\,\mathrm{d}x = \int \phi\partial_t\rho\,\mathrm{d}x$, we obtain the **Fokker–Planck** (or **forward Kolmogorov**) **equation**:

Fokker–Planck Equation

The density $\rho(t,x)$ of the SDE solution satisfies

$$\partial_t\rho = \mathcal{L}^*\rho := -\sum_i \partial_{x_i}(b^i\rho) + \tfrac{1}{2}\sum_{i,k}\partial^2_{x_ix_k}(a^{ik}\rho).$$

In one dimension with $b(x) = -\theta x$ and $a = \sigma^2$ (the OU process):

$$\partial_t\rho = \theta\partial_x(x\rho) + \tfrac{\sigma^2}{2}\partial_{xx}\rho.$$

The stationary solution is $\rho_\infty(x) \propto e^{-\theta x^2/\sigma^2}$, the Gaussian $\mathcal{N}(0, \sigma^2/(2\theta))$.

17.2 Stationary Distributions and Detailed Balance

Definition: Detailed Balance

A probability density ρ^* satisfies **detailed balance** (or is **reversible**) for the SDE if

$$b^i(x)\rho^*(x) = \tfrac{1}{2}\sum_k \partial_{x_k}(a^{ik}(x)\rho^*(x)) \quad \text{for all } i.$$

Detailed balance implies stationarity ($\mathcal{L}^*\rho^* = 0$), but not conversely. For gradient diffusions $b(x) = -\nabla U(x)$ with constant σ, the density $\rho^*(x) \propto e^{-2U(x)/\sigma^2}$ (the Gibbs measure) satisfies detailed balance.

17.3 The Kramers–Smoluchowski Equation

The **overdamped Langevin SDE** is

$$\mathrm{d}X_t = -\nabla U(X_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t,$$

describing the position of a particle in a potential U at inverse temperature β. The Fokker–Planck equation is the **Kramers–Smoluchowski** equation, and the unique stationary distribution is the Boltzmann distribution $\rho^* \propto e^{-\beta U}$.

Example: Double-well potential and metastability

Let $U(x) = (x^2-1)^2/4$. The Boltzmann measure $\rho^* \propto e^{-\beta U(x)}$ is bimodal with peaks at $x = \pm 1$, becoming sharper as $\beta \to \infty$ (low temperature). The Kramers escape rate over the barrier at $x = 0$ is approximately $e^{-\beta \Delta U}/(2\pi)$ where $\Delta U = U(0) - U(1) = 1/4$. This is Arrhenius' law.

17.4 The Probability Current and the Continuity Equation

The Fokker–Planck equation can be rewritten in a form that makes manifest the conservation of total probability and exposes the physical interpretation of the diffusion as a flow.

Definition: Probability Current

The **probability current** associated to the Fokker–Planck equation is the vector field

$$J^i(t,x) := b^i(x)\rho(t,x) - \tfrac{1}{2}\sum_k \partial_{x_k}\big(a^{ik}(x)\rho(t,x)\big)\,.$$

The Fokker–Planck equation is then equivalent to the **continuity equation**

$$\partial_t \rho + \nabla \cdot J = 0.$$

Remark: Conservation of Total Probability

Integrating the continuity equation over $\mathbb{R}^d$ and using the divergence theorem (assuming $J \to 0$ sufficiently fast at infinity) gives $\frac{\mathrm{d}}{\mathrm{d}t}\int \rho(t,x)\,\mathrm{d}x = -\int \nabla \cdot J\,\mathrm{d}x = 0$: total probability mass is conserved for all t, as it must be since $\rho(t,\cdot)$ is always a probability density. This is the diffusion-theoretic analogue of conservation of mass in fluid dynamics, with J playing the role of the mass flux; indeed, the overdamped Langevin equation of the previous section is frequently studied in the physics literature directly in this hydrodynamic language.

Current at Stationarity

The stationary density ρ^* satisfies $\nabla \cdot J^* = 0$, where J^* is the current evaluated at $\rho = \rho^*$. Two qualitatively distinct situations arise:

(i) **Equilibrium (detailed balance)**: $J^* \equiv 0$ everywhere. This is exactly the detailed balance condition of the previous section, restated as the vanishing of the current rather than merely its divergence.

(ii) **Non-equilibrium steady state**: $J^* \neq 0$ but $\nabla \cdot J^* = 0$ (a divergence-free, non-zero current). This occurs whenever b is not a gradient field, and describes systems with a persistent probability circulation even at stationarity — e.g. a particle on a ring subject to a constant non-conservative torque.

17.5 Entropy Dissipation and Convergence to Equilibrium

Definition: Relative Entropy (KL Divergence) to Equilibrium

For the Boltzmann stationary density $\rho^* \propto e^{-\beta U}$, define the relative entropy of the time-dependent density $\rho(t,\cdot)$ with respect to ρ^* as

$$H(t) := D_{\mathrm{KL}}(\rho(t,\cdot)\,\|\,\rho^*) = \int \rho(t,x)\ln\frac{\rho(t,x)}{\rho^*(x)}\,\mathrm{d}x.$$

Theorem: Monotone Entropy Dissipation

For the overdamped Langevin dynamics with detailed balance, $H(t)$ is non-increasing in t, with dissipation rate

$$-\frac{\mathrm{d}}{\mathrm{d}t}H(t) = \beta^{-1}\int \rho(t,x)\left|\nabla \ln\frac{\rho(t,x)}{\rho^*(x)}\right|^2 \mathrm{d}x =: \beta^{-1}\,I(t) \geq 0,$$

where $I(t)$ is the **relative Fisher information**. If ρ^* satisfies a log-Sobolev inequality with constant λ (Chapter 24), then $H(t) \leq H(0)\,e^{-2\lambda t/\beta}$, giving **exponential convergence to equilibrium in relative entropy**.

Remark: The Free Energy Functional and Gradient Flow Structure

The combination $F(\rho) := \beta^{-1}\int \rho \ln \rho \, \mathrm{d}x + \int U\rho \, \mathrm{d}x$ (free energy: entropy plus potential energy) decreases along the Fokker–Planck flow at exactly the rate $-\dot{F}(t) = \beta^{-1} I(t) \geq 0$ computed above, and is minimised precisely by ρ^*. This identifies the overdamped Langevin Fokker–Planck equation as the **gradient flow** of the free energy functional F with respect to the Wasserstein-2 metric on probability measures — the celebrated Jordan–Kinderlehrer–Otto (JKO) variational formulation of the heat-type equation, which reframes the entire convergence-to-equilibrium theory of this section as a statement about a curve of steepest descent in the space of measures, and which underlies modern connections between diffusion processes and optimal transport.

Chapter 18

The Backward Kolmogorov Equation and Feynman–Kac

18.1 The Backward Equation

Define the **value function** $u(s,x) := \mathbb{E}^{s,x}[g(X_T)] = \mathbb{E}[g(X_T) \mid X_s = x]$ for a bounded measurable g. Under Lipschitz conditions and smoothness of g, $u \in C^{1,2}((0,T)\times\mathbb{R}^d)$ and satisfies:

Theorem: Backward Kolmogorov Equation

$$\partial_s u + \mathcal{L}u = 0 \quad \text{on } (0,T)\times\mathbb{R}^d, \qquad u(T,x) = g(x).$$

This is the **backward parabolic PDE** associated to the generator $\mathcal{L}$. The solution is given probabilistically by $u(s,x) = P_{s,T}g(x)$.

18.2 Feynman–Kac Formula

The Feynman–Kac formula extends the above to include a potential term:

Theorem: Feynman–Kac

Let $g\colon \mathbb{R}^d \to \mathbb{R}$ be bounded continuous, $V\colon \mathbb{R}^d \to \mathbb{R}$ bounded below, and define

$$u(s,x) = \mathbb{E}^{s,x}\left[e^{-\int_s^T V(X_r)\,\mathrm{d}r} g(X_T)\right].$$

Then $u \in C^{1,2}$ and satisfies

$$\partial_s u + \mathcal{L}u - Vu = 0, \quad u(T,x) = g(x).$$

Application: Black–Scholes PDE

For GBM $\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t$ (risk-neutral measure) and payoff $g(S_T) = (S_T - K)^+$, the Feynman–Kac formula with $V = r$ gives the discounted option price:

$$C(t,s) = e^{-r(T-t)}\mathbb{E}^{t,s}[(S_T - K)^+],$$

which satisfies the **Black–Scholes PDE**:

$$\partial_t C + rs\partial_s C + \tfrac{\sigma^2 s^2}{2}\partial_{ss}^2 C - rC = 0, \quad C(T,s) = (s-K)^+.$$

18.3 Connection to the Heat Equation and Explicit Solutions

For $b \equiv 0$ and $\sigma = I$ (standard Brownian motion), the backward Kolmogorov equation reduces to the heat equation $\partial_t u = \frac{1}{2}\Delta u$. The Feynman–Kac formula then gives the probabilistic representation of solutions to the heat equation:

$$u(t,x) = \int_{\mathbb{R}^d} p(T-t,x,y)g(y)\,\mathrm{d}y,$$

where p is the heat kernel. This is a special case of a deep correspondence: every parabolic PDE with probabilistic coefficients (i.e. with a positive semi-definite diffusion matrix) has a probabilistic representation via Feynman–Kac.

18.4 Feynman–Kac with a Source Term

Theorem: Feynman–Kac with Inhomogeneous Term

Let $f\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ be bounded continuous. The function

$$u(s,x) = \mathbb{E}^{s,x}\left[e^{-\int_s^T V(X_r)\,\mathrm{d}r} g(X_T) + \int_s^T e^{-\int_s^r V(X_\ell)\,\mathrm{d}\ell} f(r,X_r)\,\mathrm{d}r\right]$$

satisfies the inhomogeneous backward equation

$$\partial_s u + \mathcal{L}u - Vu + f = 0, \qquad u(T,x) = g(x).$$

Sketch. Apply Itô's formula to $M_r := e^{-\int_s^r V(X_\ell)\,\mathrm{d}\ell}\,u(r,X_r) + \int_s^r e^{-\int_s^\ell V}\,f(\ell,X_\ell)\,\mathrm{d}\ell$ for $r \in [s,T]$. The drift of M_r vanishes precisely because u solves the stated PDE, so M_r is a local martingale; evaluating $\mathbb{E}[M_T] = M_s$ (using boundedness for uniform integrability) and substituting $u(T,\cdot) = g$ recovers the stated representation. This is the same martingale technique used to prove the homogeneous Feynman–Kac theorem above, merely tracking an additional accumulating term. □

Remark: Probabilistic Meaning of the Three Terms

The representation combines three financially and physically meaningful pieces: the terminal payoff $g(X_T)$, discounted by the multiplicative factor $\exp(-\int_s^T V)$ (interpreted as either a discount rate or, in the killed-process interpretation below, a survival probability); and a running "coupon" or source term $f(r, X_r)$, accrued continuously and likewise discounted from time r back to s. In financial applications, f represents a continuous dividend or cash-flow stream, and V the (possibly state-dependent) interest rate; in physics, f represents a source term in a reaction-diffusion equation and V a local reaction or absorption rate.

18.5 The Killed Process Interpretation

The factor $e^{-\int_s^T V(X_r)\,\mathrm{d}r}$ in the Feynman–Kac formula admits a vivid probabilistic interpretation when $V \geq 0$: it is the survival probability of a particle subject to a random, state-dependent killing rate.

Definition: Killed Diffusion

Given $V \geq 0$ and an independent exponential clock $\mathcal{E} \sim \mathrm{Exp}(1)$, define the **killing time**

$$\zeta := \inf\left\{ t \geq 0 : \int_0^t V(X_r)\ \mathrm{d}r \geq \mathcal{E} \right\}.$$

The **killed process** is X_t for $t < \zeta$ and an absorbing cemetery state $\dagger$ for $t \geq \zeta$.

Survival Probability

By the memoryless property of the exponential distribution, $\mathbb{P}(\zeta > t \mid (X_r)_{r \leq t}) = \exp\left(-\int_0^t V(X_r)\ \mathrm{d}r\right)$, so that for any bounded g,

$$\mathbb{E}\left[e^{-\int_0^T V(X_r)\,\mathrm{d}r}\, g(X_T) \right] = \mathbb{E}\left[g(X_T)\, \mathbf{1}_{\{\zeta > T\}} \right] = \mathbb{E}\left[g(\tilde{X}_T) \right],$$

where $\tilde{X}_T$ denotes X_T on the event of survival and the expectation of g at the cemetery state is taken to be 0. This recovers exactly the discounting factor of the homogeneous Feynman–Kac formula as the probability that the particle has not yet been killed by time T.

Remark: Killing and the Generator of the Sub-Markov Semigroup

The killed process is a (sub-)Markov process with generator $\mathcal{L} - V$ acting on the space of functions vanishing at $\dagger$; the loss of mass at rate V is exactly what converts the conservative semigroup (P_t) of Chapter 13 into the sub-conservative "Feynman–Kac semigroup" $(P_t^V f)(x) := \mathbb{E}^x[e^{-\int_0^t V(X_r)\,\mathrm{d}r} f(X_t)]$, whose generator is formally $\mathcal{L} - V$. This killed-process viewpoint gives an alternative,

purely probabilistic derivation of both the homogeneous and inhomogeneous Feynman–Kac formulas above, without appeal to Itô's formula, and is the natural language for first-passage and exit-time problems such as those of Chapter 19.

Chapter 19

Diffusion Processes and the Connection to Partial Differential Equations

19.1 Elliptic Equations and the Dirichlet Problem

Let $D \subset \mathbb{R}^d$ be a bounded domain, $f\colon D \to \mathbb{R}$ and $g\colon \partial D \to \mathbb{R}$ continuous. Consider the boundary value problem

$$\mathcal{L}u = -f \text{ in } D, \quad u = g \text{ on } \partial D.$$

Theorem: Probabilistic Representation (elliptic)

Under ellipticity of a and regularity of ∂D, the unique solution is

$$u(x) = \mathbb{E}^x\left[\int_0^{\tau_D} f(X_s)\,\mathrm{d}s + g(X_{\tau_D})\right],$$

where $\tau_D = \inf\{t \geq 0 : X_t \notin D\}$.

19.2 Harmonic Functions and the Maximum Principle

A function $h\colon D \to \mathbb{R}$ is **$\mathcal{L}$-harmonic** if $\mathcal{L}h = 0$ in D. By Dynkin's formula, $h(X_{t\wedge\tau_D})$ is then a martingale. The maximum principle follows immediately: a non-constant $\mathcal{L}$-harmonic function attains its maximum only on ∂D.

19.3 Harnack Inequality and Regularity

Harnack Inequality

Let a be uniformly elliptic (i.e. $\lambda|\xi|^2 \le \sum_{ik} a^{ik}\xi^i\xi^k \le \Lambda|\xi|^2$ for $0 < \lambda \le \Lambda$). For any compact $K \subset D$ there exists $C = C(K, D, \lambda, \Lambda)$ such that for any non-negative $\mathcal{L}$-harmonic h on D,

$$\sup_K h \le C \inf_K h.$$

19.4 Mean Exit Times and the Poisson Equation

Expected Exit Time

Let τ_D be the first exit time of X from a bounded domain D. Then $u(x) := \mathbb{E}^x[\tau_D]$ solves the Poisson equation

$$\mathcal{L}u = -1 \text{ in } D, \qquad u = 0 \text{ on } \partial D,$$

a special case of the representation theorem above with $f \equiv 1$, $g \equiv 0$.

Example: Exit Time from an Interval

For Brownian motion on $\mathbb{R}$ started at $x \in (-a, a)$ with $\mathcal{L} = \frac{1}{2}\partial_{xx}$, the equation $\frac{1}{2}u'' = -1$ with $u(\pm a) = 0$ has the explicit solution $u(x) = a^2 - x^2$, recovering the classical formula $\mathbb{E}^x[\tau_{(-a,a)}] = a^2 - x^2$ derived via martingale methods (optional stopping applied to $B_t^2 - t$) in Volume I-style exercises and revisited in Chapter 1's Exercise 1 of this volume in the related first-passage setting.

19.5 The Principal Eigenvalue and Exponential Exit Asymptotics

Definition: Principal Dirichlet Eigenvalue

For uniformly elliptic $\mathcal{L}$ on a bounded domain D with smooth boundary, the **principal Dirichlet eigenvalue** $\lambda_1 = \lambda_1(D) > 0$ is the smallest λ for which there exists $\phi > 0$ in D, $\phi = 0$ on ∂D, solving

$$-\mathcal{L}\phi = \lambda\phi \text{ in } D, \qquad \phi = 0 \text{ on } \partial D.$$

Theorem: Exponential Tail of the Exit Time

The principal eigenvalue governs the exponential decay rate of the survival probability:

$$\mathbb{P}^x(\tau_D > t) = \phi(x)\, e^{-\lambda_1 t}\,(1 + o(1)) \quad \text{as } t \to \infty,$$

uniformly for x in compact subsets of D; equivalently $-\lim_{t\to\infty} \frac{1}{t} \ln \mathbb{P}^x(\tau_D > t) = \lambda_1$.

Sketch. Write $u(t,x) := \mathbb{P}^x(\tau_D > t) = \mathbb{E}^x[\mathbf{1}_{\{\tau_D > t\}}]$. By the same martingale argument as in the Kolmogorov backward equation (Chapter 13), u solves $\partial_t u = \mathcal{L}u$ in D with $u(t,\cdot) = 0$ on ∂D and $u(0,\cdot) = 1$. Expanding $u(t,\cdot)$ in the (complete, orthogonal in $L^2(D)$) eigenbasis $\{\phi_k\}$ of $-\mathcal{L}$ with eigenvalues $0 < \lambda_1 < \lambda_2 \le \cdots$ gives $u(t,x) = \sum_k c_k \phi_k(x) e^{-\lambda_k t}$; as $t \to \infty$ the $k = 1$ term dominates exponentially, giving the stated asymptotic with $\phi = \phi_1$. □

Remark: Connection to Large Deviations

The principal eigenvalue $\lambda_1(D)$ is the diffusion-theoretic analogue of the large-deviation rate functions studied for Markov chains in Volume I: both quantify the exponential rate at which a process avoids an unlikely set (here, remaining inside D rather than exiting). As D shrinks, $\lambda_1(D) \to \infty$ in a manner governed by the geometry of D (e.g. for a ball of radius r in $\mathbb{R}^d$ under standard Brownian motion, λ_1 scales like r^{-2}), a fact that is the continuum counterpart of the spectral gap estimates revisited for general diffusions in Chapter 24.

Chapter 20

Stochastic Optimal Control

20.1 The Stochastic Control Problem

Stochastic optimal control combines the SDE theory of Part III with an optimization problem over the drift, extending Markov decision process theory (Volume I, Chapter 23) to the diffusion setting.

Definition: Controlled SDE

A **controlled SDE** on $\mathbb{R}^d$ is

$$\mathrm{d}X_t^u = b(t, X_t^u, u_t)\,\mathrm{d}t + \sigma(t, X_t^u)\,\mathrm{d}W_t, \quad X_0^u = x,$$

where $(u_t)_{t\geq 0}$ is the **control process**, taking values in a compact set $\mathcal{U} \subseteq \mathbb{R}^k$. The goal is to minimise the **cost functional**

$$J(u; x) = \mathbb{E}^x\left[\int_0^T \ell(t, X_t^u, u_t)\,\mathrm{d}t + g(X_T^u)\right]$$

over all admissible controls, where ℓ is the running cost and g the terminal cost.

Definition: Value Function

The **value function** of the control problem is

$$V(t, x) = \inf_{u \in \mathcal{U}} J(u; t, x) = \inf_{u} \mathbb{E}^{t,x}\left[\int_t^T \ell(s, X_s^u, u_s)\,\mathrm{d}s + g(X_T^u)\right].$$

20.2 The Hamilton–Jacobi–Bellman Equation

Theorem: Hamilton–Jacobi–Bellman (HJB) Equation

Under appropriate regularity conditions, the value function $V \in C^{1,2}([0,T)\times\mathbb{R}^d)$ satisfies the **HJB equation**:

$$-\partial_t V = \inf_{u\in\mathcal{U}} \left\{ \ell(t,x,u) + \sum_i b^i(t,x,u)\partial_{x_i} V + \tfrac{1}{2}\sum_{ik} a^{ik}(t,x)\partial^2_{x_i x_k} V \right\},$$

with terminal condition $V(T,x) = g(x)$.

Proof sketch via dynamic programming. The **dynamic programming principle** (DPP) asserts that for any stopping time $\tau \le T$:

$$V(t,x) = \inf_u \mathbb{E}^{t,x}\left[\int_t^\tau \ell(s, X_s^u, u_s)\, \mathrm{d}s + V(\tau, X_\tau^u) \right].$$

Applying Itô's formula to $V(\tau, X_\tau^u)$ and using the DPP, sending $\tau \to t^+$ yields the HJB equation in the limit. □

Corollary: Optimal Control via HJB

If $V \in C^{1,2}$ solves the HJB equation, the optimal control at (t,x) is the minimiser

$$u^*(t,x) = \arg\min_{u\in\mathcal{U}} \left\{ \ell(t,x,u) + \sum_i b^i(t,x,u)\partial_{x_i} V(t,x) \right\},$$

and the optimal controlled process X^{u^*} satisfies $J(u^*;t,x) = V(t,x)$.

20.3 The Linear-Quadratic Regulator

The **linear-quadratic regulator** (LQR) is the canonical exactly solvable stochastic control problem, and the foundation of Kalman filtering.

Example: LQR Problem

Controlled linear dynamics: $\mathrm{d}X_t = (AX_t + Bu_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$ on $\mathbb{R}^d$. Quadratic cost:

$$J(u) = \mathbb{E}\left[\int_0^T \tfrac{1}{2}(X_t^\top Q X_t + u_t^\top R u_t)\,\mathrm{d}t + \tfrac{1}{2} X_T^\top P_T X_T \right],$$

with Q, R, P_T positive semi-definite and R positive definite. The HJB equation admits a quadratic ansatz $V(t,x) = \tfrac{1}{2}x^\top P(t) x + c(t)$ where $P(t)$ satisfies the **matrix Riccati ODE**:

$$-\dot{P} = A^\top P + PA - PBR^{-1}B^\top P + Q, \quad P(T) = P_T.$$

The optimal control is $u^*(t,x) = -R^{-1}B^\top P(t)x$, a linear feedback law. The constant term satisfies $\dot{c}(t) = -\frac{1}{2}\operatorname{tr}(\sigma^\top P(t)\sigma)$, capturing the irreducible cost due to noise.

20.4 Connection to Backward SDEs

The HJB equation has a probabilistic counterpart via **backward stochastic differential equations** (BSDEs), which express the value function as the solution of a stochastic equation rather than a PDE.

Definition: Backward SDE

A **BSDE** with terminal condition $\xi \in L^2(\mathcal{F}_T)$ and driver $f : [0,T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is the pair (Y_t, Z_t) satisfying

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)\, \mathrm{d}s - \int_t^T Z_s\, \mathrm{d}W_s, \quad t \in [0,T].$$

Theorem: Pardoux–Peng Existence and Uniqueness

If $\xi \in L^2(\mathcal{F}_T)$ and f is Lipschitz in (y,z) uniformly in t, then the BSDE has a unique solution $(Y,Z) \in \mathcal{S}^2 \times \mathcal{H}^2$ where $\mathcal{S}^2$ denotes square-integrable adapted processes with continuous paths.

The connection to control is as follows: for a nonlinear cost functional, the value function $V(t,x)$ corresponds to $Y_t = V(t,X_t)$ and $Z_t = \nabla_x V(t,X_t)\cdot\sigma(t,X_t)$ where (Y,Z) solves a BSDE with driver $f(t,y,z) = -\ell(t,X_t,u^*(t,X_t)) - \mathcal{H}(t,X_t,z)$ (the Hamiltonian). BSDEs thus provide a purely probabilistic alternative to the PDE approach via HJB, and are particularly useful when V lacks the $C^{1,2}$ regularity required for the classical HJB theory.

Example: BSDE representation of European option price

For a European option with payoff $g(S_T)$ under a diffusion model $\mathrm{d}S_t = \mu(S_t)\,\mathrm{d}t + \sigma(S_t)\,\mathrm{d}W_t$, the arbitrage-free price $Y_t = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[g(S_T) \mid \mathcal{F}_t]$ solves a linear BSDE:

$$Y_t = g(S_T) + \int_t^T rY_s\, \mathrm{d}s - \int_t^T Z_s\, \mathrm{d}W_s,$$

where $Z_t = e^{-r(T-t)}\sigma(S_t)\partial_s u(t,S_t)$ is the **delta hedge** (the hedging strategy). The Black–Scholes formula is thus the explicit solution of this linear BSDE.

Part V

Advanced Topics

Chapter 21

Stratonovich Calculus

21.1 Motivation: SDEs in Differential Geometry

When studying SDEs on manifolds, the Itô integral is inconvenient because it does not transform correctly under smooth changes of coordinates. The **Stratonovich integral**, denoted $\int H_s \circ \mathrm{d}W_s$, satisfies the classical chain rule and is therefore the natural integral for geometric applications.

Definition: Stratonovich Integral

The **Stratonovich integral** of $H \in \mathcal{H}^2$ against Brownian motion is defined by midpoint discretisation:

$$\int_0^T H_s \circ \mathrm{d}W_s := \lim_{\|\Pi\|\to 0} \sum_i \tfrac{1}{2}(H_{t_i} + H_{t_{i+1}})(W_{t_{i+1}} - W_{t_i}),$$

or equivalently in terms of the Itô integral:

$$\int_0^T H_s \circ \mathrm{d}W_s = \int_0^T H_s \,\mathrm{d}W_s + \tfrac{1}{2}[H, W]_T.$$

Theorem: Stratonovich Chain Rule

Let $f \in C^2$ and $X_t = X_0 + \int_0^t b_s \,\mathrm{d}s + \int_0^t \sigma_s \circ \mathrm{d}W_s$ (Stratonovich SDE). Then

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) \,\mathrm{d}X_s \quad \text{(Stratonovich integral)},$$

which is the classical chain rule, without the Itô correction term.

21.2 Conversion Between Itô and Stratonovich

Any Stratonovich SDE $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t) \circ \mathrm{d}W_t$ can be rewritten as an Itô SDE:

$$\mathrm{d}X_t = \big(b(X_t) + \tfrac{1}{2}\sigma'(X_t)\sigma(X_t)\big)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t.$$

The extra drift $\frac{1}{2}\sigma'(X_t)\sigma(X_t)$ is the **Itô–Stratonovich correction**.

Example: Geometric Brownian Motion

The Stratonovich SDE $\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma S_t \circ \mathrm{d}W_t$ converts to the Itô SDE $\mathrm{d}S_t = (\mu + \frac{\sigma^2}{2})S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t$, with solution $S_t = S_0 e^{\mu t + \sigma W_t}$. This differs from the Itô GBM $\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t$, which has solution $S_t = S_0 e^{(\mu - \sigma^2/2)t + \sigma W_t}$.

21.3 SDEs on Manifolds

Let M be a smooth Riemannian manifold. A diffusion on M is described by a Stratonovich SDE

$$\mathrm{d}X_t = V_0(X_t)\,\mathrm{d}t + \sum_{j=1}^{m} V_j(X_t) \circ \mathrm{d}W_t^j,$$

where $V_0, V_1, \ldots, V_m$ are smooth vector fields on M. The generator of this diffusion is $\mathcal{L} = V_0 + \frac{1}{2}\sum_j V_j^2$ (in the Lie-algebra sense). The **Hörmander condition** states that the Lie algebra generated by $\{V_j, j = 0, \ldots, m\}$ spans the tangent space at each point; under this condition, $\mathcal{L}$ is hypoelliptic and the diffusion has a smooth density everywhere.

21.4 The Wong–Zakai Approximation Theorem

The Stratonovich integral's compatibility with the ordinary chain rule is not merely a formal convenience: it is exactly the integral that arises as the limit of solutions to *ordinary* differential equations driven by smooth approximations of Brownian motion, which gives Stratonovich calculus its physical legitimacy as the correct model for real (band-limited) noise.

Theorem: Wong–Zakai

Let W^n be piecewise-linear interpolations of Brownian motion on a mesh of size $1/n$ (or, more generally, any sequence of smooth approximations converging to W uniformly on compacts with $W^n \to W$ a.s.). Let X_t^n solve the ordinary (pathwise) differential equation

$$\dot{X}_t^n = b(X_t^n) + \sigma(X_t^n)\,\dot{W}_t^n, \qquad X_0^n = x_0.$$

Then $X_t^n \to X_t$ a.s. and uniformly on compacts, where X_t solves the **Stratonovich** SDE

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t) \circ \mathrm{d}W_t.$$

Remark: The Physical Significance of the Wong–Zakai Theorem

The Wong–Zakai theorem resolves an apparent paradox: physical noise is never truly white (it has finite correlation time and smooth, differentiable sample paths), so equations of motion perturbed by "real" noise are ordinary differential equations in the classical sense, to which the usual chain rule applies without correction. As the correlation time of the physical noise shrinks to zero — the idealisation that produces white noise $\dot{W}_t$ in the limit — the *limiting* stochastic equation is the **Stratonovich**, not the Itô, equation. This is why the Itô–Stratonovich correction term $\frac{1}{2}\sigma'\sigma$ is often called a "noise-induced drift": it has no analogue in the underlying smooth approximation and emerges purely from the singular limit. Whenever a stochastic model is meant to approximate a physical system with band-limited noise, the Stratonovich formulation, converted to Itô form via Section 21.2 if needed, is generally the more physically faithful starting point.

Example: Brownian Motion on the Circle

Consider Brownian motion constrained to the unit circle $S^1 \subset \mathbb{R}^2$, described by the Stratonovich SDE for the angle θ_t,

$$\mathrm{d}\theta_t = \circ\, \mathrm{d}W_t,$$

i.e. $\theta_t = \theta_0 + W_t \pmod{2\pi}$. Because the Stratonovich integral obeys the ordinary chain rule, the embedded process $X_t = (\cos\theta_t, \sin\theta_t)$ satisfies, by direct (Stratonovich) differentiation,

$$\mathrm{d}X_t = \begin{pmatrix} -\sin\theta_t \\ \cos\theta_t \end{pmatrix} \circ\, \mathrm{d}W_t,$$

manifestly tangent to the circle at every instant — a geometric constraint (staying on the manifold) that the corresponding Itô formulation would only satisfy after including a compensating radial drift, since the naïve Itô SDE for $(\cos\theta_t, \sin\theta_t)$ with the same diffusion coefficient would drift off the circle. This illustrates concretely why the Stratonovich formulation, not Itô, is the natural language for diffusions constrained to a manifold.

Chapter 22

Local Time and the Tanaka Formula

22.1 Local Time of Brownian Motion

Definition: Local Time

The **local time** L_t^a of Brownian motion at level $a \in \mathbb{R}$ is the density of occupation time:

$$\int_0^t \mathbf{1}_{\{B_s \in A\}}\,\mathrm{d}s = \int_A L_t^a\,\mathrm{d}a \quad \text{a.s. for all Borel } A.$$

Intuitively, $L_t^a\,\mathrm{d}a$ is the (rescaled) amount of time Brownian motion spends near level a up to time t.

Theorem: Tanaka Formula

For Brownian motion and $a \in \mathbb{R}$,

$$|B_t - a| = |B_0 - a| + \int_0^t \operatorname{sgn}(B_s - a)\,\mathrm{d}W_s + L_t^a,$$

where $\operatorname{sgn}(x) = \mathbf{1}_{x>0} - \mathbf{1}_{x\le 0}$.

The Tanaka formula is the Itô formula for the non-smooth function $f(x) = |x-a|$. The local time L_t^a plays the role of the second derivative term, now a singular measure.

22.2 The Occupation Times Formula and Generalised Itô–Tanaka

Itô–Tanaka–Meyer Formula

Let $f\colon \mathbb{R} \to \mathbb{R}$ be the difference of two convex functions with left-continuous derivative f' and second derivative measure μ_f. For a continuous semimartin-

gale X,

$$f(X_t) = f(X_0) + \int_0^t f'(X_{s-})\,\mathrm{d}X_s + \tfrac{1}{2}\int_{\mathbb{R}} L_t^a \mu_f(\mathrm{d}a).$$

Example: Reflected Brownian motion

Taking $f(x) = x^+$, whose derivative is $\mathbf{1}_{x>0}$ and second derivative is the point mass at 0:

$$B_t^+ = \int_0^t \mathbf{1}_{B_s>0}\,\mathrm{d}W_s + \tfrac{1}{2}L_t^0.$$

Hence $L_t^0 = 2\left(B_t^+ - \int_0^t \mathbf{1}_{B_s>0}\,\mathrm{d}W_s\right)$.

22.3 Reflected and Absorbed Brownian Motion

Definition: Reflected Brownian Motion

Reflected Brownian motion (RBM) on $[0,\infty)$ is the process $|B_t|$, which satisfies

$$|B_t| = |B_0| + \int_0^t \operatorname{sgn}(B_s)\,\mathrm{d}W_s + L_t^0.$$

Alternatively, RBM is the unique strong solution to $\mathrm{d}R_t = \mathrm{d}W_t + \mathrm{d}L_t$ where $R_t \geq 0$ and L increases only when $R_t = 0$ (the boundary local time).

RBM on bounded intervals plays a central role in queueing theory (as the diffusion limit of the $M/M/1$ queue) and in mathematical finance (barrier options).

22.4 Joint Continuity and the Ray–Knight Theorems

So far L_t^a has been fixed at a single level a. A deeper structural fact is that local time can be realised jointly, as a continuous random field in (t,a), and that this field has a remarkable Markovian structure in the spatial variable when time is replaced by an inverse local time.

Theorem: Trotter's Joint Continuity

There is a version of $(L_t^a)_{t\geq 0, a\in\mathbb{R}}$ that is a.s. jointly continuous in (t,a).

Definition: Inverse Local Time

The **inverse local time at** 0 is $\tau_\ell := \inf\{t \geq 0 : L_t^0 > \ell\}$, $\ell \geq 0$. By the strong Markov property and the strict increase of L^0 off the zero set of B, $(\tau_\ell)_{\ell\geq 0}$ is a stable subordinator of index $\tfrac{1}{2}$: $\mathbb{E}[e^{-\lambda\tau_\ell}] = e^{-\ell\sqrt{2\lambda}}$.

Theorem: First Ray–Knight Theorem

Let $\tau_1 = \inf\{t : B_t = 1\}$ (the first hitting time of level 1 for Brownian motion started at 0), and consider the local time field $\{L^a_{\tau_1} : 0 \le a \le 1\}$ stopped at this hitting time. Then, as a process in the spatial variable a,

$$\left(L^{1-a}_{\tau_1}\right)_{0\le a\le 1} \overset{d}{=} \left(Z^2_a\right)_{0\le a\le 1},$$

where $(Z_a)_{0\le a\le 1}$ is the radial part of a 2-dimensional Brownian motion started at 0, i.e. Z^2 is a squared Bessel process of dimension 2 (Chapter 4) run for "time" a.

Remark: Local Time as a Diffusion in Space

The Ray–Knight theorems are among the most striking results in the theory of Brownian motion: they assert that, viewed as a function of the *spatial* variable rather than time, local time is itself governed by a diffusion (here, a squared Bessel process). This spatial Markov property has no a priori reason to hold and is special to Brownian motion (and, more generally, to the class of processes amenable to excursion theory below). Ray–Knight theory underlies sharp results on the structure of Brownian local time near its maximum and is a key tool in the analysis of self-intersection local times and in superprocess theory, both beyond the scope of this volume.

22.5 Itô's Excursion Theory

Definition: Excursion Process

An **excursion** of Brownian motion away from 0 is a piece of path $(B_s)_{s\in[g,d]}$ between two consecutive zeros $g < d$ of B with $B_s \neq 0$ for $s \in (g, d)$. Parametrising excursions by the inverse local time $\ell = L^0_g = L^0_d$, the **excursion process** $(e_\ell)_{\ell\ge 0}$, $e_\ell(u) := B_{\tau_{\ell^-}+u}$ for $0 \le u \le \tau_\ell - \tau_{\ell^-}$, records the shape of each excursion as a function of ℓ.

Theorem: Itô's Excursion Theorem

The excursion process $(e_\ell)_{\ell\ge 0}$ is a Poisson point process on the space of excursion paths, with intensity measure $\mathrm{d}\ell \otimes n(\,\mathrm{d}e)$ for a fixed (σ-finite, infinite total mass) measure n on excursion space called the **Itô excursion measure**.

Remark: Excursion Theory as a Unifying Framework

Itô excursion theory recasts Brownian motion as a juxtaposition of independent, identically structured excursions glued together at the (random, Lebesgue-null) zero set, governed by the Poissonian intensity n. This perspective gives a clean, conceptual re-derivation of many results scattered through this volume: the

arc-sine law (Exercise 3) follows from the scaling property of n; the reflection principle (Chapter 3) is a direct consequence of the symmetry of n under $e \mapsto -e$; and the Ray–Knight theorems above can be re-derived as statements about the Lévy measure of the subordinator (τ_ℓ) associated to n. Excursion theory generalises naturally to any Markov process with a recurrent point, and is a standard tool in the modern theory of Markov processes.

Chapter 23

Ergodic Theory for Diffusions

23.1 Invariant Measures for SDEs

A probability measure π on $\mathbb{R}^d$ is **invariant** for the SDE $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$ if

$$\int \mathcal{L}f \,\mathrm{d}\pi = 0 \quad \text{for all } f \in C_c^\infty(\mathbb{R}^d).$$

This is equivalent to $\mathcal{L}^*\pi = 0$ (in the distributional sense), i.e. π is a stationary solution of the Fokker–Planck equation.

Theorem: Existence of Invariant Measure (Khasminskii–Bogoliubov–Krylov)

If the SDE solution is non-explosive and satisfies a *tightness* condition (e.g. there exists V with $\mathcal{L}V \le -cV + C$), then an invariant probability measure exists.

23.2 Ergodicity and Mixing

Theorem: Ergodic Theorem for Diffusions

If (X_t) is ergodic with unique invariant measure π, then for any $f \in L^1(\pi)$,

$$\frac{1}{T}\int_0^T f(X_s)\,\mathrm{d}s \xrightarrow{T\to\infty} \int f\,\mathrm{d}\pi \quad \mathbb{P}\text{-a.s.}$$

Example: Underdamped Langevin dynamics

The **kinetic Langevin SDE** (or underdamped Langevin) describes position q and velocity v of a particle in a potential U:

$$\begin{aligned} \mathrm{d}q_t &= v_t\,\mathrm{d}t, \\ \mathrm{d}v_t &= -\gamma v_t\,\mathrm{d}t - \nabla U(q_t)\,\mathrm{d}t + \sqrt{2\gamma\beta^{-1}}\,\mathrm{d}W_t, \end{aligned}$$

with $\gamma > 0$ the friction coefficient. The unique invariant measure is proportional to $e^{-\beta H(q,v)}$ where $H = |v|^2/2 + U(q)$ is the Hamiltonian. This SDE is used extensively in molecular dynamics and MCMC sampling (Langevin Monte Carlo).

23.3 The Spectral Gap and Poincaré Inequality

Theorem: Spectral Gap and L^2 Decay

If the generator $\mathcal{L}$ has a spectral gap $\lambda > 0$ (i.e. $-\mathcal{L} \geq \lambda$ on $L^2_0(\pi)$, the zero-mean functions), then

$$\|P_t f - \pi f\|_{L^2(\pi)} \leq e^{-\lambda t} \|f - \pi f\|_{L^2(\pi)}.$$

The spectral gap of the OU process with $\mathcal{L} = -\theta x \partial_x + \frac{\sigma^2}{2}\partial_{xx}$ is $\lambda = \theta$.

Poincaré Inequality

The spectral gap $\lambda > 0$ is equivalent to the **Poincaré inequality**: for all $f \in H^1(\pi)$,

$$\mathrm{Var}_\pi(f) \leq \frac{1}{\lambda} \int |\nabla f|^2 \, \mathrm{d}\pi.$$

A stronger condition, the **log-Sobolev inequality** $\mathrm{Ent}_\pi(f^2) \leq \frac{2}{\rho}\int |\nabla f|^2 \, \mathrm{d}\pi$, implies both the Poincaré inequality and exponential decay of relative entropy, and is the key tool for sharp mixing time bounds for continuous-time diffusions analogous to the cutoff results of Volume I.

23.4 The Bakry–Émery Criterion

Verifying a spectral gap or log-Sobolev inequality directly from the definition is typically infeasible. The Bakry–Émery Γ_2 calculus gives a purely local, differential sufficient condition — a curvature-type bound on the generator — that implies both.

Definition: Carré du Champ and Bakry–Émery Curvature

For the generator $\mathcal{L} = \frac{1}{2}\Delta - \nabla U \cdot \nabla$ of the overdamped Langevin diffusion with potential U (Chapter 17), the **carré du champ** operator is $\Gamma(f) := \frac{1}{2}\left[\mathcal{L}(f^2) - 2f\mathcal{L}f\right] = |\nabla f|^2$, and the iterated carré du champ is

$$\Gamma_2(f) := \tfrac{1}{2}\left[\mathcal{L}\Gamma(f) - 2\Gamma(f, \mathcal{L}f)\right] = \left\|\nabla^2 f\right\|_{HS}^2 + \langle \nabla^2 U \, \nabla f, \nabla f\rangle.$$

The diffusion satisfies the $\mathrm{CD}(\rho, \infty)$ **curvature-dimension condition** if $\Gamma_2(f) \geq \rho\,\Gamma(f)$ for all smooth f, which holds if and only if $\nabla^2 U \geq \rho\, I$ (i.e. U is ρ-strongly convex).

Theorem: Bakry–Émery Implies Log-Sobolev

If CD(ρ, ∞) holds with $\rho > 0$, then the invariant measure $\pi \propto e^{-2U}$ satisfies the log-Sobolev inequality with constant ρ, and hence the Poincaré inequality with spectral gap $\lambda \geq \rho$.

Example: Gaussian Measure and the Harmonic Potential

For $U(x) = \frac{\theta}{2}|x|^2$ (so that π is Gaussian, $\mathcal{N}(0, I/(2\theta))$, the invariant measure of the OU process), $\nabla^2 U = \theta I$, so CD(θ, ∞) holds with $\rho = \theta$. The Bakry–Émery criterion then gives spectral gap $\lambda \geq \theta$, matching exactly the value $\lambda = \theta$ computed directly from the spectrum of the OU generator in the theorem above — confirming that the Bakry–Émery bound is sharp in the Gaussian case.

Remark: Curvature in a Probabilistic Sense

The terminology "curvature-dimension condition" is not incidental: the Bakry–Émery criterion is the Eulerian (analytic) counterpart of a genuine Ricci curvature lower bound when the diffusion lives on a Riemannian manifold rather than flat $\mathbb{R}^d$ (the Lichnerowicz–Obata framework), and CD(ρ, ∞) with $\rho > 0$ is the flat-space avatar of positive Ricci curvature. This dictionary between convexity of a potential and curvature of a diffusion underlies the modern theory of optimal transport and concentration of measure, of which the Poincaré and log-Sobolev inequalities above are foundational instances.

23.5 Uniqueness of the Invariant Measure and Strong Feller Diffusions

Existence of an invariant measure (Theorem 24.2) does not by itself guarantee uniqueness, nor the convergence stated in the ergodic theorem above; both require an irreducibility-type condition ruling out the diffusion splitting into non-communicating invariant pieces.

Definition: Strong Feller Property

The semigroup (P_t) is **strong Feller** if $P_t f$ is continuous for every bounded measurable f (not merely for continuous f, as in the ordinary Feller property of Chapter 13).

Theorem: Doob's Uniqueness Criterion

If (P_t) is strong Feller and **topologically irreducible** (for every open $U \neq \emptyset$ and x, $P_t(x, U) > 0$ for some $t > 0$), then (P_t) has at most one invariant probability measure, and if it exists, $\|P_t(x, \cdot) - \pi\|_{TV} \to 0$ for every starting point x (not merely π-a.e. x).

Remark: Hypoellipticity Supplies the Strong Feller Property

For diffusions satisfying Hörmander's condition (Chapter 14), the transition kernel $p_t(x, y)$ is smooth in both variables for $t > 0$, which immediately implies the strong Feller property: $P_t f(x) = \int f(y) p_t(x, y)\, \mathrm{d}y$ depends continuously on x even for merely bounded f, because the dependence is mediated by a smooth kernel. This is one of the main practical routes to verifying uniqueness and convergence to equilibrium for degenerate diffusions such as the kinetic Langevin equation of Section 24.2, where the noise acts only on the velocity coordinate v and not directly on q, yet Hörmander's bracket condition $[V_1, V_0] \neq 0$ (with V_0 the drift and V_1 the noise vector field) restores hypoellipticity and hence uniqueness of the Gibbs invariant measure.

Chapter 24

Rough Paths: An Introduction

24.1 Motivation

The theory of rough paths, introduced by Terry Lyons (1998), extends the notion of a controlled differential equation to paths of very low regularity (Hölder exponent less than $1/2$), including fractional Brownian motion with Hurst parameter $H < 1/2$ and other non-semimartingale drivers.

24.2 Young Integration and its Limits

For Hölder paths $f \in C^\alpha$ and $g \in C^\beta$ with $\alpha + \beta > 1$, the **Young integral** $\int f \, \mathrm{d}g$ is well-defined as a limit of Riemann sums. The condition $\alpha + \beta > 1$ fails for Brownian motion ($\alpha < 1/2$ and $\beta < 1/2$), which is why the Itô theory is needed.

For paths rougher than Brownian motion ($\alpha < 1/4$), even the Itô theory fails, and rough paths provide the correct framework.

24.3 The Rough Path

Definition: Rough Path

A **rough path** above a path $X \in C^\alpha([0,T];V)$ for $\alpha \in (1/3, 1/2]$ is a pair $\mathbf{X} = (X, \mathbb{X})$ where $\mathbb{X}\colon [0,T]^2 \to V \otimes V$ provides the "iterated integral" $\mathbb{X}_{s,t} \approx \int_s^t (X_r - X_s) \otimes \mathrm{d}X_r$, satisfying Chen's relation $\mathbb{X}_{s,t} = \mathbb{X}_{s,u} + \mathbb{X}_{u,t} + (X_u - X_s) \otimes (X_t - X_u)$.

Lyons' Continuity Theorem

The solution map $\mathbf{X} \mapsto Y$ of the rough differential equation $\mathrm{d}Y_t = f(Y_t)\,\mathrm{d}\mathbf{X}_t$ is locally Lipschitz in the rough path metric. In particular, it extends continuously from smooth paths to rough paths.

This continuity is the key feature distinguishing rough paths from the Itô theory,

where the solution map is only continuous in a much weaker (probabilistic) sense.

24.4 Brownian Motion as a Rough Path

The canonical rough path lift of Brownian motion is obtained by setting $\mathbb{B}^{ij}_{s,t} = \int_s^t (B^i_r - B^i_s)\,\mathrm{d}W^j_r$ (Itô integral). The resulting rough path $\mathbf{B}^{\text{Itô}} = (B, \mathbb{B})$ defines the Itô calculus. The Stratonovich lift instead sets $\mathbb{B}^{ij}_{s,t} = \int_s^t (B^i_r - B^i_s) \circ \mathrm{d}W^j_r$; these differ by $\frac{1}{2}\delta_{ij}(t-s)$, which is precisely the Itô–Stratonovich correction.

Rough paths theory thus provides a unified framework in which both the Itô and Stratonovich calculi arise as different choices of the iterated integral enhancement.

24.5 The Signature and Chen's Relation

Definition: The Signature of a Path

For a smooth path $X\colon [0,T] \to V$, the **signature** is the sequence of all iterated integrals,

$$S(X)_{0,T} = \left(1,\ \int_0^T \mathrm{d}X_{t_1},\ \int_{0<t_1<t_2<T} \mathrm{d}X_{t_1} \otimes \mathrm{d}X_{t_2},\ \ldots\right) \in T((V)) = \bigoplus_{n\geq 0} V^{\otimes n},$$

an element of the (formal) tensor algebra over V. Truncating at level 2 recovers exactly the pair $(X, \mathbb{X})$ of the rough path definition above.

Chen's Relation as a Group Law

The signature satisfies $S(X)_{s,t} = S(X)_{s,u} \otimes S(X)_{u,t}$ (the shuffle/concatenation identity), making $\{S(X)_{s,t} : s \leq t\}$ into a path in the group of group-like elements of $T((V))$ under the tensor product. This is precisely Chen's relation, here elevated from the level-2 truncation used to define a rough path to the full, infinite-level signature.

Remark: The Signature as a Universal, Faithful Summary of a Path

A foundational result of rough path theory (Chen, and later Hambly–Lyons) shows that the signature determines a path up to reparametrisation and "tree-like" excursions: two paths have the same signature if and only if one can be obtained from the other by inserting or removing such excursions. This has made the signature a popular feature map in modern machine learning for sequential data (time series, handwriting, financial price paths): the (truncated) signature provides a reparametrisation-invariant, algebraically structured summary that captures the full nonlinear interaction between coordinates, generalising the simple summary statistics (mean, total variation) familiar from elementary calculus.

24.6 Rough Differential Equations and the Davie Expansion

Theorem: Davie's Definition of Solutions to RDEs

A controlled process Y is a solution to the rough differential equation $\mathrm{d}Y_t = f(Y_t)\ \mathrm{d}\mathbf{X}_t$ if, locally,

$$Y_t - Y_s = f(Y_s)(X_t - X_s) + Df(Y_s)f(Y_s) : \mathbb{X}_{s,t} + o(|t-s|^{2\alpha})$$

as $t \to s$, where $\mathbb{X}_{s,t}$ supplies the second-order (Lévy area type) correction needed because $\alpha \leq 1/2$ makes the first-order Euler approximation $f(Y_s)(X_t - X_s)$ alone insufficiently accurate.

Remark: Connection to the Stochastic Taylor Expansion

The Davie expansion is the deterministic, pathwise analogue of the stochastic Taylor expansion used to derive the Milstein scheme in Chapter 15: both retain the first-order term $f(Y_s)\Delta X$ and a second-order correction built from an iterated integral of X against itself (the Lévy area $\mathbb{X}_{s,t}$, equivalently the bracket increment $[X]_{s,t}$ in the semimartingale case). The rough path perspective explains *why* the Milstein correction is unavoidable for accurate strong approximation: it is the same algebraic term that appears, deterministically, in the definition of the solution itself, rather than merely an artefact of a particular numerical discretisation.

Appendix A

Review of Measure Theory and Conditional Expectation

A.1 Key Results

We collect here the measure-theoretic results used throughout the volume without proof.

Monotone Convergence Theorem. If $0 \le X_n \nearrow X$ a.s., then $\mathbb{E}[X_n] \nearrow \mathbb{E}[X]$.

Dominated Convergence Theorem. If $X_n \to X$ a.s. and $|X_n| \le Y$ with $\mathbb{E}[Y] < \infty$, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

Fubini–Tonelli. For (σ-finite) product measures, $\int \int f \, \mathrm{d}\mu \, \mathrm{d}\nu = \int \int f \, \mathrm{d}\nu \, \mathrm{d}\mu$ when $f \ge 0$ or $f \in L^1$.

A.2 Conditional Expectation

For $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and a sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, the conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ is the unique (a.s.) $\mathcal{G}$-measurable Y satisfying $\mathbb{E}[Y \mathbf{1}_A] = \mathbb{E}[X \mathbf{1}_A]$ for all $A \in \mathcal{G}$.

Key properties:

(i) **Tower property**: $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[X \mid \mathcal{H}]$ for $\mathcal{H} \subseteq \mathcal{G}$;

(ii) **Pull-out**: $\mathbb{E}[YX \mid \mathcal{G}] = Y \mathbb{E}[X \mid \mathcal{G}]$ when Y is $\mathcal{G}$-measurable and $YX \in L^1$;

(iii) **Jensen's inequality**: $\varphi(\mathbb{E}[X \mid \mathcal{G}]) \le \mathbb{E}[\varphi(X) \mid \mathcal{G}]$ for convex φ;

(iv) **Independence**: If X is independent of $\mathcal{G}$, then $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$.

A.3 The Borel–Cantelli Lemmas

First Borel–Cantelli Lemma

If $\{A_n\}$ are events with $\sum_n \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n \text{ occurs for infinitely many } n) = 0$.

Second Borel–Cantelli Lemma

If $\{A_n\}$ are *independent* events with $\sum_n \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_n \text{ occurs for infinitely many } n) = 1$.

These lemmas are the basic engine behind almost every almost-sure statement in this volume that is proved by a discretisation-and-summation argument: the law of the iterated logarithm (Chapter 3), the non-differentiability of Brownian paths (Chapter 3), and the a.s. convergence rates underlying the Euler–Maruyama analysis (Chapter 15) all reduce, at some point in the proof, to checking a summability condition and invoking the first Borel–Cantelli lemma.

A.4 Versions and Regularity: The Kolmogorov Continuity Theorem

A subtlety used repeatedly without comment in the main text is the distinction between a stochastic process and a particular *version* of it (two processes that agree a.s. at every fixed time, but whose joint path regularity may differ); the construction of Brownian motion in Chapter 3, and the various smoothness statements made about it, all implicitly rely on selecting a good version.

Definition: Version (Modification) of a Process

Two processes (X_t) and $(\tilde{X}_t)$ are **versions** (or modifications) of each other if $\mathbb{P}(X_t = \tilde{X}_t) = 1$ for every fixed t. They need not have the same sample paths, even though all finite-dimensional distributions agree.

Theorem: Kolmogorov Continuity Criterion

Suppose $(X_t)_{t\in[0,T]}$ satisfies, for some $\alpha, \beta, C > 0$ and all $s, t \in [0, T]$,

$$\mathbb{E}[|X_t - X_s|^{\alpha}] \le C\,|t - s|^{1+\beta}\,.$$

Then X has a version with sample paths that are a.s. Hölder continuous of every exponent $\gamma < \beta/\alpha$.

Remark: The Workhorse Behind Path Regularity Throughout This Volume

This single theorem, applied with $\alpha = 2n$ and the explicit Gaussian moment bound $\mathbb{E}[|B_t - B_s|^{2n}] = (2n-1)!!\ |t-s|^n$, is exactly what justifies the claim (used freely from Chapter 3 onward) that Brownian motion has a continuous version, and more precisely a version that is Hölder continuous of every exponent $< 1/2$ — letting $n \to \infty$ in the criterion above pushes $\beta/\alpha = (n-1)/(2n) \to 1/2$. The same criterion, applied to increments of fractional Brownian motion (Chapter 5) with the appropriate moment bound, yields the Hölder-$(H-\varepsilon)$ regularity asserted there, and underlies the construction of continuous modifications of solutions to SDEs in Chapter 12.

Appendix B

Gaussian Processes

B.1 Gaussian Random Vectors

A random vector $\mathbf{X} = (X_1, \dots, X_n)$ is **Gaussian** if every linear combination $\sum_i a_i X_i$ is a normal random variable. It is characterised by its mean vector $\mu = \mathbb{E}[\mathbf{X}]$ and covariance matrix $\Sigma_{ij} = \operatorname{Cov}(X_i, X_j)$. The characteristic function is $\mathbb{E}[e^{i\langle \theta, \mathbf{X}\rangle}] = e^{i\langle\theta,\mu\rangle - \frac{1}{2}\langle\theta,\Sigma\theta\rangle}$.

B.2 Gaussian Hilbert Spaces

A **Gaussian Hilbert space** is a closed subspace $H \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$ consisting of centred Gaussian random variables. The Itô integral $\{I(h) := \int_0^T h(t)\,\mathrm{d}W_t : h \in L^2([0,T])\}$ is a Gaussian Hilbert space, and the map $h \mapsto I(h)$ is an isometry from $L^2([0,T])$ onto this space.

B.3 Reproducing Kernel Hilbert Spaces

The covariance function $K(s,t) = \mathbb{E}[X_s X_t]$ of a Gaussian process is a positive-definite kernel. The associated **reproducing kernel Hilbert space** (RKHS) $\mathcal{H}_K$ consists of functions $f\colon [0,T] \to \mathbb{R}$ that can be represented as $f(t) = \mathbb{E}[X_t Z]$ for some $Z \in L^2$. For Brownian motion, $K(s,t) = \min(s,t)$ and the RKHS is the Cameron–Martin space $H^1_0([0,T])$ of absolutely continuous functions with square-integrable derivative and $f(0) = 0$.

B.4 Wick Products and Hermite Polynomials

The decomposition of L^2 functionals of Brownian motion into orthogonal pieces — already glimpsed in the Itô chaos representation of Chapter 6 — has a particularly explicit form on a Gaussian Hilbert space, expressed through Hermite polynomials.

Definition: Hermite Polynomials

The (probabilist's) **Hermite polynomials** H_n are defined by $H_0 = 1$ and the recursion $H_{n+1}(x) = xH_n(x) - nH_{n-1}(x)$, or equivalently by the generating function $e^{tx-t^2/2} = \sum_{n\geq 0} \frac{t^n}{n!} H_n(x)$. They form an orthogonal basis of $L^2(\mathbb{R}, \gamma)$ where γ is the standard normal density, with $\mathbb{E}[H_n(Z)H_m(Z)] = n!\,\delta_{nm}$ for $Z \sim \mathcal{N}(0,1)$.

Definition: Wick Power

For $Z \sim \mathcal{N}(0, \sigma^2)$ in a Gaussian Hilbert space H, the **n-th Wick power** is $Z^{:n:} := \sigma^n H_n(Z/\sigma)$, the orthogonal projection of Z^n onto the n-th Wiener chaos. Wick powers are characterised by $\mathbb{E}[Z^{:n:}] = 0$ for $n \geq 1$ and orthogonality across different orders: $\mathbb{E}[Z^{:n:} W^{:m:}] = 0$ for $n \neq m$ whenever Z, W lie in the same Gaussian Hilbert space.

Wick Powers of the Itô Integral and the Chaos Decomposition

For $I(h) = \int_0^T h_s \, \mathrm{d}W_s$ with $\|h\|_{L^2} = 1$, the iterated Itô integral satisfies

$$n! \int_{0<t_1<\cdots<t_n<T} h_{t_1} \cdots h_{t_n} \, \mathrm{d}W_{t_1} \cdots \mathrm{d}W_{t_n} = I(h)^{:n:} = H_n(I(h)),$$

giving an explicit bridge between the abstract Itô–Wiener chaos decomposition of Chapter 6 and concrete Hermite polynomials of a single Gaussian variable in the case of a one-dimensional driving kernel h.

Remark: Why Wick Calculus Matters

Wick products provide the correct notion of "multiplication" for generalised functionals of white noise (such as those arising in stochastic PDEs driven by space-time white noise), where ordinary pointwise products of distributions are ill-defined. Replacing ordinary products with Wick products systematically removes the divergent self-interaction terms that would otherwise appear — precisely the renormalisation procedure underlying the modern theory of singular SPDEs (regularity structures, Hairer 2014), of which the rough path theory of Chapter 25 is a finite-dimensional precursor.

B.5 Conditioning of Gaussian Processes

Theorem: Gaussian Conditioning Formula

Let (X, Y) be jointly Gaussian with $X \in \mathbb{R}^n$, $Y \in \mathbb{R}^m$, mean (μ_X, μ_Y) and covariance blocks $\Sigma_{XX}, \Sigma_{XY}, \Sigma_{YY}$ (with Σ_{YY} invertible). Then the conditional law of X given $Y = y$ is Gaussian with

$$\mathbb{E}[X \mid Y = y] = \mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y - \mu_Y), \qquad \mathrm{Cov}(X \mid Y = y) = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX},$$

independent of the conditioning value y (a special feature of the Gaussian case).

Example: The Brownian Bridge Revisited

Conditioning Brownian motion $(B_t)_{t\in[0,1]}$ on $B_1 = 0$ recovers the Brownian bridge of Chapter 5: applying the formula with $X = B_t$, $Y = B_1$, $\Sigma_{XY} = \min(t,1) = t$, $\Sigma_{YY} = 1$ gives $\mathbb{E}[B_t \mid B_1 = 0] = 0$ and $\operatorname{Cov}(B_s, B_t \mid B_1 = 0) = \min(s,t) - st$, exactly the bridge covariance $s(1-t)$ for $s \le t$ stated in Chapter 5, providing a second, purely linear-algebraic derivation alongside the SDE-based construction given there.

Remark: Gaussian Process Regression

The conditioning formula above is the mathematical core of **Gaussian process regression**, a widely used nonparametric Bayesian method: given noisy observations $y_i = f(t_i) + \varepsilon_i$ of an unknown function modelled as a centred Gaussian process with covariance kernel K, the posterior mean and covariance of f at a new point t^* are given by exactly the formula above, with Σ_{XY}, Σ_{YY} built from K evaluated at the observed and query points (plus observation noise variance on the diagonal of Σ_{YY}). The reproducing kernel Hilbert space of K, introduced earlier in this chapter, characterises precisely which functions can be approximated with vanishing posterior variance as the number of observations grows.

Appendix C

Notation and Symbol Table

Symbol	Meaning	First use
$(\Omega, \mathcal{F}, \mathbb{P})$	Probability space	Ch. 1
$(\mathcal{F}_t)_{t\geq 0}$	Filtration	Ch. 1
$\mathcal{F}_\tau$	Stopped σ-algebra at τ	Ch. 1
(B_t), (W_t)	Standard Brownian motion	Ch. 3
$[X]_t$, $[X,Y]_t$	Quadratic variation, covariation	Ch. 2, 7
$\langle M\rangle_t$	Predictable quadratic variation	Ch. 2
$\mathcal{L}$	Generator of SDE	Ch. 11
$a = \sigma\sigma^\top$	Diffusion matrix	Ch. 11
$\rho(t,x)$	Density of X_t	Ch. 17
$\mathcal{L}^*$	Fokker–Planck operator	Ch. 17
L_t^a	Local time at level a	Ch. 22
$\mathcal{E}(X)_t$	Doléans–Dade exponential	Ch. 7
$\mathcal{H}^2$	Space of square-integrable integrands	Ch. 6
$\mathbf{D}^{1,2}$	Malliavin Sobolev space	Ch. 15
$D_s F$	Malliavin derivative	Ch. 15
δ	Skorokhod integral	Ch. 15
Φ	Standard normal CDF	Ch. 3
τ_D	First exit time from D	Ch. 1
P_t, $P_{s,t}$	Semigroup, transition kernel	Ch. 11
π	Invariant (stationary) measure	Ch. 24
$\mathbf{X}$	Rough path	Ch. 25

Appendix D

Worked Applications

This chapter develops five self-contained applications of the theory, each carried through from model formulation to explicit quantitative conclusion. The aim is to show the machinery of stochastic calculus in use on problems of genuine scientific or financial interest, rather than as isolated technique. Each application draws on several chapters simultaneously: the reader will see Itô's formula, Girsanov's theorem, the Fokker–Planck equation, and existence–uniqueness results cooperate in a single argument.

D.1 European Option Pricing via Girsanov and the Black–Scholes Formula

The model

Fix a finite horizon $T > 0$ and a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ carrying a standard Brownian motion (W_t). The market consists of a riskless bond $B_t = e^{rt}$ and a risky asset with dynamics

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}W_t, \qquad S_0 > 0,$$

where $\mu \in \mathbb{R}$ is the (physical) drift and $\sigma > 0$ the volatility. A European call option with strike K and maturity T pays $(S_T - K)^+$ at time T.

Step 1: Solution of the GBM SDE

Applying Itô's formula to $\log S_t$ with $f(x) = \log x$, $f'(x) = 1/x$, $f''(x) = -1/x^2$:

$$\mathrm{d}(\log S_t) = \frac{1}{S_t}\,\mathrm{d}S_t - \frac{1}{2}\cdot\frac{1}{S_t^2}\,(\mathrm{d}S_t)^2 = \left(\mu - \tfrac{\sigma^2}{2}\right)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t.$$

Integrating:

$$S_t = S_0 \exp\!\left[\left(\mu - \tfrac{\sigma^2}{2}\right)t + \sigma W_t\right].$$

Step 2: Change of measure to the risk-neutral world

Define the **market price of risk** $\theta := (\mu - r)/\sigma$ and the Doléans exponential

$$Z_T = \exp\left(-\theta W_T - \tfrac{1}{2}\theta^2 T\right).$$

By Itô's formula, $\mathrm{d}Z_t = -\theta Z_t \,\mathrm{d}W_t$, so $Z_T > 0$ and $\mathbb{E}[Z_T] = 1$ (Novikov's condition holds since θ is constant). Define the **risk-neutral measure** $\mathbb{Q}$ on $\mathcal{F}_T$ by

$$\left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_T} = Z_T.$$

By Girsanov's theorem (Chapter 9), the process $\widetilde{W}_t := W_t + \theta t$ is a standard $\mathbb{Q}$-Brownian motion. Rewriting the dynamics:

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}W_t = \mu S_t \,\mathrm{d}t + \sigma S_t(\,\mathrm{d}\widetilde{W}_t - \theta\,\mathrm{d}t) = r S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}\widetilde{W}_t,$$

so under $\mathbb{Q}$ the discounted price $\tilde{S}_t = e^{-rt}S_t$ is a martingale, as required by no-arbitrage.

Step 3: Risk-neutral pricing

The no-arbitrage price of the call at time 0 is the discounted $\mathbb{Q}$-expectation of the payoff:

$$C_0 = e^{-rT}\mathbb{E}^{\mathbb{Q}}\left[(S_T - K)^+\right].$$

Under $\mathbb{Q}$,

$$S_T = S_0 \exp\left[\left(r - \tfrac{\sigma^2}{2}\right)T + \sigma\widetilde{W}_T\right],$$

and $\widetilde{W}_T \sim \mathcal{N}(0, T)$ under $\mathbb{Q}$. Writing $\widetilde{W}_T = \sqrt{T}\,Z$ with $Z \sim \mathcal{N}(0,1)$:

$$C_0 = e^{-rT}\int_{-\infty}^{\infty}\left(S_0 e^{(r-\sigma^2/2)T + \sigma\sqrt{T}\,z} - K\right)^+ \frac{e^{-z^2/2}}{\sqrt{2\pi}}\,\mathrm{d}z.$$

The integrand is positive when $z > d_2 := \left(\log(K/S_0) - (r - \sigma^2/2)T\right)/(\sigma\sqrt{T})$. A direct Gaussian computation gives the **Black–Scholes formula**:

Black–Scholes Formula

$$C_0 = S_0\,\Phi(d_1) - K e^{-rT}\Phi(d_2),$$

where Φ is the standard normal CDF and

$$d_{1,2} = \frac{\log(S_0/K) + (r \pm \sigma^2/2)T}{\sigma\sqrt{T}}.$$

Proof. Split the integral at $z = d_2$. For the term involving S_0, complete the square in the exponent:

$$\begin{aligned}
&\int_{d_2}^{\infty} S_0 e^{(r-\sigma^2/2)T + \sigma\sqrt{T}\,z}\,\frac{e^{-z^2/2}}{\sqrt{2\pi}}\,\mathrm{d}z \\
&= S_0 e^{rT}\int_{d_2}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\tfrac{1}{2}(z - \sigma\sqrt{T})^2\right)\mathrm{d}z = S_0 e^{rT}\,\Phi(-d_2 + \sigma\sqrt{T}) = S_0 e^{rT}\,\Phi(d_1).
\end{aligned}$$

The term involving K is $K\,\mathbb{P}^{\mathbb{Q}}(\widetilde{W}_T/\sqrt{T} > d_2) = K\,\Phi(-d_2)$. Multiplying through by e^{-rT} and noting $-d_2 = \log(S_0/K)/(\sigma\sqrt{T}) + (r - \sigma^2/2)\sqrt{T}/\sigma$ gives $\Phi(-d_2) = \Phi(d_2)$ after sign flip in the definition, completing the computation. $\square$

Step 4: The hedging strategy

By the martingale representation theorem (Chapter 10), the $\mathbb{Q}$-martingale $M_t = \mathbb{E}^{\mathbb{Q}}[e^{-rT}(S_T - K)^+ \mid \mathcal{F}_t]$ can be written as $M_t = C_0 + \int_0^t \Delta_s \,\mathrm{d}\tilde{S}_s$ for a unique process $(\Delta_t) \in \mathcal{H}^2$. Computing $M_t = e^{-rt}C(t, S_t)$ where $C(t,s) = s\Phi(d_1(t,s)) - Ke^{-r(T-t)}\Phi(d_2(t,s))$ and applying Itô's formula shows that the **delta** of the option is

$$\Delta_t = \frac{\partial C}{\partial s}(t, S_t) = \Phi(d_1(t, S_t)),$$

the unique self-financing replicating strategy. The key observation is that $C(t,s)$ satisfies the **Black–Scholes PDE**

$$\partial_t C + rs\,\partial_s C + \tfrac{1}{2}\sigma^2 s^2\,\partial_{ss} C = rC,$$

which is precisely the Feynman–Kac representation (Chapter 18) of the discounted expectation above.

D.2 The Ornstein–Uhlenbeck Process: Full Probabilistic Analysis

The model

The Ornstein–Uhlenbeck (OU) process satisfies

$$\mathrm{d}X_t = -\theta X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \qquad X_0 = x_0 \in \mathbb{R},$$

with $\theta > 0$, $\sigma > 0$. It is the canonical mean-reverting diffusion, modelling interest rates (Vasicek), stochastic volatility, and the velocity of a particle in a viscous fluid.

Step 1: Explicit solution via Itô's product rule

Multiply both sides by the integrating factor $e^{\theta t}$ and apply Itô's product rule to $Y_t := e^{\theta t}X_t$:

$$\mathrm{d}Y_t = \theta e^{\theta t}X_t\,\mathrm{d}t + e^{\theta t}\,\mathrm{d}X_t = \theta e^{\theta t}X_t\,\mathrm{d}t + e^{\theta t}(-\theta X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t) = \sigma e^{\theta t}\,\mathrm{d}W_t.$$

Integrating from 0 to t:

$$Y_t = x_0 + \sigma\int_0^t e^{\theta s}\,\mathrm{d}W_s,$$

so

$$X_t = x_0 e^{-\theta t} + \sigma\int_0^t e^{-\theta(t-s)}\,\mathrm{d}W_s. \tag{D.1}$$

Step 2: Moments and covariance structure

Since the stochastic integral in (D.1) has mean zero:

$$\mathbb{E}[X_t] = x_0 e^{-\theta t} \longrightarrow 0 \quad \text{as } t \to \infty.$$

By the Itô isometry:

$$\operatorname{Var}(X_t) = \sigma^2 \int_0^t e^{-2\theta(t-s)}\, \mathrm{d}s = \frac{\sigma^2}{2\theta}\left(1 - e^{-2\theta t}\right) \longrightarrow \frac{\sigma^2}{2\theta}.$$

For $s \le t$, the covariance follows from (D.1) and independence of increments:

$$\begin{aligned}\operatorname{Cov}(X_s, X_t) &= \sigma^2 \int_0^s e^{-\theta(s-u)} e^{-\theta(t-u)}\, \mathrm{d}u = \sigma^2 e^{-\theta(s+t)} \int_0^s e^{2\theta u}\, \mathrm{d}u \\ &= \frac{\sigma^2}{2\theta}\left(e^{-\theta(t-s)} - e^{-\theta(t+s)}\right).\end{aligned}$$

Step 3: Stationarity and the invariant measure

Since (X_t) is a Gaussian process (the stochastic integral is Gaussian), it is fully characterised by its mean and covariance. As $t \to \infty$, the distribution converges to $\mathcal{N}(0, \sigma^2/(2\theta))$, which is the unique invariant measure $\pi = \mathcal{N}(0, \sigma^2/(2\theta))$. Direct verification via the Fokker–Planck equation (Chapter 17): the stationary density ρ^* satisfies

$$\partial_x(\theta x\, \rho^*) + \tfrac{1}{2}\sigma^2 \partial_{xx}\rho^* = 0.$$

Substituting $\rho^*(x) = C\exp(-\theta x^2/\sigma^2)$ (a Gaussian with variance $\sigma^2/(2\theta)$):

$$\theta\rho^* + \theta x(-2\theta x/\sigma^2)\rho^* + \tfrac{\sigma^2}{2}\cdot\left(\frac{-2\theta}{\sigma^2} + \frac{4\theta^2 x^2}{\sigma^4}\right)\rho^* = \left(\theta - \frac{2\theta^2 x^2}{\sigma^2} - \theta + \frac{2\theta^2 x^2}{\sigma^2}\right)\rho^* = 0. \checkmark$$

Step 4: Exponential ergodicity and the spectral gap

The generator of the OU process is $\mathcal{L} = -\theta x\, \partial_x + \frac{\sigma^2}{2}\partial_{xx}$. On $L^2(\pi)$, the eigenvalue equation $\mathcal{L}f = -\lambda f$ is solved by Hermite polynomials $H_n(x\sqrt{2\theta}/\sigma)$: the eigenvalues are $\lambda_n = n\theta$ for $n = 0, 1, 2, \ldots$. The spectral gap is $\lambda_1 = \theta$, meaning

$$\left\| P_t f - \pi(f) \right\|_{L^2(\pi)} \le e^{-\theta t} \left\| f - \pi(f) \right\|_{L^2(\pi)},$$

so the semigroup contracts at the mean-reversion rate θ, in exact agreement with the decay of $\mathbb{E}[X_t]$.

D.3 Barrier Option Pricing via the Reflection Principle

The model

In the Black–Scholes market of Section 1, consider a **down-and-out call**: an option that pays $(S_T - K)^+$ at time T if the asset price never fell below a barrier $L < \min(S_0, K)$, and pays zero otherwise. With $L < S_0$, the relevant stopping time is

$$\tau_L := \inf\{t \ge 0 : S_t \le L\}.$$

The payoff is $(S_T - K)^+ \mathbf{1}_{\{\tau_L > T\}}$.

Step 1: Reduction to Brownian motion

Under $\mathbb{Q}$, $S_t = S_0 \exp(r - \sigma^2/2)t + \sigma\widetilde{W}_t)$, so $S_t \le L$ iff $\widetilde{W}_t \le b$ where

$$b := \frac{\log(L/S_0)}{\sigma} < 0, \qquad \widetilde{W}_t^* := \widetilde{W}_t + \frac{r - \sigma^2/2}{\sigma}\, t.$$

More precisely, let $\nu := (r - \sigma^2/2)/\sigma$ and define the drifted Brownian motion $\hat{W}_t := \widetilde{W}_t + \nu t$. Then $S_t = S_0 e^{\sigma \hat{W}_t}$ and $\tau_L = \inf\{t : \hat{W}_t \le b\}$. The price becomes

$$C_0^{\text{DOC}} = e^{-rT}\mathbb{E}^{\mathbb{Q}}\left[(S_0 e^{\sigma \hat{W}_T} - K)^+ \mathbf{1}_{\{\min_{s \le T} \hat{W}_s > b\}}\right].$$

Step 2: The reflection principle for drifted Brownian motion

For a Brownian motion $(\hat{W}_t)$ with constant drift ν and a level $b < 0$, the joint law of $(\hat{W}_T, \min_{s \le T} \hat{W}_s)$ is computed via Girsanov: first for $\nu = 0$ the reflection principle (Chapter 3) gives

$$\mathbb{P}\left(\min_{s \le T} B_s \le b,\ B_T \ge x\right) = \mathbb{P}(B_T \le 2b - x) \quad \text{for } x > b,$$

since the path reflected at the first hitting time of b has the same distribution as a path reaching $2b - x$ directly. For the drifted case, Girsanov's theorem with drift ν introduces a density $e^{\nu \hat{W}_T - \nu^2 T/2}$. One computes the joint density of $(\hat{W}_T, \tau_b)$ and integrates.

Step 3: Explicit formula

After carrying out the Gaussian integration (which mirrors the Black–Scholes computation of Section 1, applied twice), the price of the down-and-out call is:

Down-and-Out Call Formula

$$C_0^{\text{DOC}} = C_0^{\text{BS}} - \left(\frac{L}{S_0}\right)^{2r/\sigma^2 - 1}\left[S_0\left(\frac{L}{S_0}\right)^2 \Phi(\tilde{d}_1) - K e^{-rT}\Phi(\tilde{d}_2)\right],$$

where C_0^{BS} is the standard Black–Scholes call price from Section 1, and

$$\tilde{d}_{1,2} = \frac{\log(L^2/(S_0 K)) + (r \pm \sigma^2/2)T}{\sigma\sqrt{T}}.$$

Remark: The Two-Term Structure

The formula has a transparent structure: the first term is the unrestricted Black–Scholes price, and the second term is the subtracted value of a *reflected* option, originating from paths that hit the barrier. The exponent $2r/\sigma^2 - 1$ is the ratio

of drift to diffusion in the Girsanov weight; for $r = \sigma^2/2$ it reduces to 1 and the formula simplifies to a pure difference. The same method extends to up-and-out options, double-barrier options, and lookback options (which depend on the running maximum or minimum of S), all via the reflection principle or its multidimensional generalisation.

D.4 Overdamped Langevin Dynamics and the Convergence to Equilibrium

The model

Let $U : \mathbb{R} \to \mathbb{R}$ be a smooth confining potential (meaning $U(x) \to +\infty$ as $|x| \to \infty$, sufficiently fast). The **overdamped Langevin SDE** is

$$\mathrm{d}X_t = -\nabla U(X_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t, \qquad X_0 = x_0,$$

where $\beta > 0$ is the inverse temperature. This is the canonical model for the motion of a particle in a potential well subject to thermal noise, and is the basis of Markov chain Monte Carlo sampling (MCMC) at temperature β^{-1}.

Step 1: Existence of the invariant measure

The Fokker–Planck equation (Chapter 17) for the density $\rho(t, x)$ of X_t is

$$\partial_t \rho = \nabla \cdot \big(\nabla U \cdot \rho + \beta^{-1} \nabla \rho\big) = \nabla \cdot (\rho\, \nabla U) + \beta^{-1} \Delta \rho.$$

We seek a stationary solution ρ^* with $\partial_t \rho^* = 0$. The **detailed balance** (or reversibility) condition reads: the probability current $J = -\rho^* \nabla U - \beta^{-1} \nabla \rho^*$ vanishes identically. Setting $J = 0$ gives

$$\frac{\nabla \rho^*}{\rho^*} = -\beta \nabla U \implies \rho^*(x) = Z^{-1} e^{-\beta U(x)},$$

the **Gibbs–Boltzmann measure** with partition function $Z = \int e^{-\beta U(x)}\,\mathrm{d}x < \infty$ (finite by the confining assumption).

Proposition: Verification of Stationarity

$\rho^*(x) = Z^{-1} e^{-\beta U(x)}$ is the unique probability density satisfying $\nabla \cdot J = 0$ with $J = 0$, and is therefore the unique invariant measure of the overdamped Langevin dynamics.

Proof. Direct substitution: $\nabla \rho^* = -\beta (\nabla U) \rho^*$, so $J = -\rho^* \nabla U - \beta^{-1}(-\beta \nabla U\, \rho^*) = 0$. Uniqueness follows from the irreducibility of the diffusion on $\mathbb{R}^d$ (the diffusion matrix $\sigma = \sqrt{2\beta^{-1}} I$ is non-degenerate) together with the fact that any invariant measure must satisfy the same current equation. □

Step 2: Log-Sobolev inequality and exponential convergence

The rate of convergence to ρ^* is controlled by the **log-Sobolev inequality** (LSI): if U satisfies $\nabla^2 U \geq \kappa I$ for some $\kappa > 0$ (strong convexity), then for all probability densities ρ with $\int \rho \log \rho \, \mathrm{d}\pi < \infty$,

$$\mathrm{Ent}_\pi(\rho) := \int \frac{\rho}{\rho^*} \log \frac{\rho}{\rho^*} \, \rho^* \, \mathrm{d}x \leq \frac{1}{2\kappa\beta^{-1}} \int \left| \nabla \log \frac{\rho}{\rho^*} \right|^2 \rho \, \mathrm{d}x.$$

This is the Bakry–Émery criterion for the LSI constant $\alpha_{\mathrm{LSI}} = \kappa\beta^{-1}$. The **entropy production identity** (de Bruijn's identity in this context) states that along the Fokker–Planck flow,

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Ent}_\pi(\rho_t) = -2\beta^{-1} \int \left| \nabla \log \frac{\rho_t}{\rho^*} \right|^2 \rho_t \, \mathrm{d}x.$$

Combining with the LSI:

Exponential Convergence to Equilibrium

If $\nabla^2 U \geq \kappa I$ with $\kappa > 0$, then

$$\mathrm{Ent}_\pi(\rho_t) \leq e^{-2\kappa t} \, \mathrm{Ent}_\pi(\rho_0).$$

In particular, the relative entropy converges to zero at rate 2κ, and by Pinsker's inequality the total variation distance satisfies

$$\|\rho_t - \rho^*\|_{\mathrm{TV}} \leq \sqrt{\tfrac{1}{2} \mathrm{Ent}_\pi(\rho_0)} \, e^{-\kappa t}.$$

Step 3: Worked example — the double-well potential

Let $U(x) = x^4/4 - x^2/2$, the standard double-well with minima at $x = \pm 1$ and local maximum at $x = 0$. Here $U''(x) = 3x^2 - 1$, which is negative at $x = 0$, so strong convexity fails globally and the LSI argument above does not apply directly. The invariant measure

$$\rho^*(x) \propto \exp\left(-\beta\left(\tfrac{x^4}{4} - \tfrac{x^2}{2}\right)\right)$$

is bimodal for large β (low temperature), with mass concentrated near ± 1. The transition rate between the two wells is governed by the **Kramers–Eyring formula**:

$$k_{\pm} \approx \frac{\sqrt{|U''(0)| \cdot U''(\pm 1)}}{2\pi} \, e^{-\beta(U(0) - U(\pm 1))} = \frac{\sqrt{1 \cdot 2}}{2\pi} \, e^{-\beta/4},$$

where $U(0) = 0$ and $U(\pm 1) = -1/4$ give the barrier height $\Delta U = 1/4$. This exponential factor $e^{-\beta/4}$ is precisely the large-deviation cost (Chapter 17) of a fluctuation crossing the potential barrier, and it drives the logarithmic divergence of the mixing time as $\beta \to \infty$.

D.5 The Euler–Maruyama Scheme: Strong Convergence with Explicit Constants

The setup

Consider the SDE on $[0, T]$:

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \qquad X_0 = x_0,$$

where $b, \sigma \colon \mathbb{R} \to \mathbb{R}$ are Lipschitz with constant $L > 0$:

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \le L|x - y| \quad \forall\, x, y \in \mathbb{R}.$$

The **Euler–Maruyama** (EM) scheme with step $h = T/N$ produces the approximation

$$X^h_{k+1} = X^h_k + b(X^h_k)\,h + \sigma(X^h_k)\,(W_{(k+1)h} - W_{kh}), \quad X^h_0 = x_0.$$

Step 1: Continuous-time interpolation

Define the continuous interpolation $\bar{X}_t := X^h_{\lfloor t/h \rfloor}$ (piecewise constant) and the EM process

$$X^h_t := x_0 + \int_0^t b(\bar{X}_s)\,\mathrm{d}s + \int_0^t \sigma(\bar{X}_s)\,\mathrm{d}W_s.$$

At the grid points $t = kh$ this coincides with the EM scheme. The error $e_t := X_t - X^h_t$ satisfies

$$e_t = \int_0^t \big[b(X_s) - b(\bar{X}_s)\big]\,\mathrm{d}s + \int_0^t \big[\sigma(X_s) - \sigma(\bar{X}_s)\big]\,\mathrm{d}W_s.$$

Step 2: The L^2 bound via Itô's isometry and Gronwall

Taking $L^2(\mathbb{P})$ norms and using $|a + b|^2 \le 2(a^2 + b^2)$:

$$\begin{aligned}\mathbb{E}[e_t^2] &\le 2T \int_0^t \mathbb{E}\big[(b(X_s) - b(\bar{X}_s))^2\big]\,\mathrm{d}s + 2\int_0^t \mathbb{E}\big[(\sigma(X_s) - \sigma(\bar{X}_s))^2\big]\,\mathrm{d}s \\ &\le 2(T+1)L^2 \int_0^t \mathbb{E}\big[(X_s - \bar{X}_s)^2\big]\,\mathrm{d}s.\end{aligned}$$

Now $X_s - \bar{X}_s = e_s + (X^h_s - \bar{X}_s)$, and $X^h_s - \bar{X}_s$ is the increment of the EM process within the current step $[s_k, s]$ where $s_k = h\lfloor s/h \rfloor$:

$$X^h_s - \bar{X}_s = \int_{s_k}^s b(\bar{X}_u)\,\mathrm{d}u + \int_{s_k}^s \sigma(\bar{X}_u)\,\mathrm{d}W_u.$$

Using the linear growth bound $|b(x)| + |\sigma(x)| \le L(1 + |x|)$ and the Itô isometry:

$$\mathbb{E}\big[(X^h_s - \bar{X}_s)^2\big] \le 2(h^2 + h)\,L^2(1 + \mathbb{E}[\bar{X}_s^2]) \le C_1 h,$$

for a constant $C_1 = C_1(L, T, x_0)$ depending on the L^2 bound on the solution (which is $O(e^{CT})$ by the standard estimate). Therefore:

$$\mathbb{E}[e_t^2] \le 2(T+1)L^2 \int_0^t \big(2\mathbb{E}[e_s^2] + C_1 h\big)\,\mathrm{d}s = 4(T+1)L^2 \int_0^t \mathbb{E}[e_s^2]\,\mathrm{d}s + 2(T+1)L^2 C_1 h t.$$

By **Gronwall's inequality** ($\phi(t) \le \alpha t + \beta \int_0^t \phi(s)\,\mathrm{d}s \implies \phi(t) \le \alpha t e^{\beta t}$) with $\alpha = 2(T+1)L^2 C_1 h$ and $\beta = 4(T+1)L^2$:

Strong Convergence of Euler–Maruyama

Under Lipschitz and linear growth conditions, the EM scheme satisfies

$$\sup_{0 \le k \le N} \mathbb{E}\big[(X_{kh} - X_{kh}^h)^2\big] \le C\,h,$$

where $C = 2(T+1)L^2 C_1 T\, e^{4(T+1)L^2 T}$. In particular, the strong L^2 error is $O(h^{1/2})$.

Step 3: Sharpness and higher-order schemes

The order $h^{1/2}$ is optimal for the EM scheme applied to SDEs with non-constant diffusion coefficient: the error comes from approximating $\sigma(X_s)$ by $\sigma(X_{s_k})$ on each step, and the Itô isometry converts this $O(h)$ coefficient error into an $O(h^{1/2})$ path error. The **Milstein scheme**, which adds the term

$$\tfrac{1}{2}\sigma(X_k^h)\sigma'(X_k^h)\big[(W_{(k+1)h} - W_{kh})^2 - h\big]$$

to the EM update, corrects the leading error term and achieves $O(h)$ strong order, at the cost of requiring σ' (in one dimension). In multiple dimensions, the Milstein correction involves iterated Itô integrals $\int_{t_k}^{t_{k+1}} \int_{t_k}^{s} \mathrm{d}W_u^i\,\mathrm{d}W_s^j$ for $i \ne j$, which are expensive to simulate and motivate higher-order methods such as the stochastic Runge–Kutta family.

Remark: Weak vs. Strong Convergence

The strong convergence rate $O(h^{1/2})$ concerns pathwise approximation: $\mathbb{E}[|X_T - X_T^h|^2]^{1/2} = O(h^{1/2})$. The **weak convergence rate** $O(h)$ of the EM scheme concerns approximation of expectations: $|\mathbb{E}[f(X_T)] - \mathbb{E}[f(X_T^h)]| = O(h)$ for smooth test functions f, a result that requires the Kolmogorov backward equation and is proved via the Feynman–Kac formula. For Monte Carlo estimation of $\mathbb{E}[f(X_T)]$, weak convergence is the relevant criterion; for strong approximation (e.g. in filtering or path-dependent derivatives), strong convergence is required.

Bibliography

[1] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., Springer, 1991.

[2] B. Øksendal, *Stochastic Differential Equations*, 6th ed., Springer, 2003.

[3] P. Protter, *Stochastic Integration and Differential Equations*, 2nd ed., Springer, 2004.

[4] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, 3rd ed., Springer, 1999.

[5] K. Itô, *On stochastic differential equations*, Mem. Amer. Math. Soc. 4 (1951).

[6] P. Lévy, *Processus stochastiques et mouvement brownien*, Gauthier-Villars, 1948.

[7] I. V. Girsanov, *On transforming a certain class of stochastic processes by absolutely continuous substitution of measures*, Theor. Probab. Appl. 5 (1960), 285–301.

[8] R. F. Bass, *Stochastic Processes*, CUP, 2011.

[9] F. C. Klebaner, *Introduction to Stochastic Calculus with Applications*, 3rd ed., Imperial College Press, 2012.

[10] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, 2nd ed., North-Holland, 1989.

[11] H. Kunita, *Stochastic Flows and Stochastic Differential Equations*, CUP, 1990.

[12] S. Kusuoka and D. Stroock, *Applications of the Malliavin calculus*, Parts I–III, J. Fac. Sci. Univ. Tokyo (1982–87).

[13] D. Nualart, *The Malliavin Calculus and Related Topics*, 2nd ed., Springer, 2006.

[14] R. Khasminskii, *Stochastic Stability of Differential Equations*, 2nd ed., Springer, 2012.

[15] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, 1992.

[16] T. Lyons, *Differential equations driven by rough signals*, Rev. Mat. Iberoam. 14 (1998), 215–310.

[17] M. Hairer, *A theory of regularity structures*, Invent. Math. 198 (2014), 269–504.

[18] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, J. Polit. Econ. 81 (1973), 637–654.

[19] R. C. Merton, *Theory of rational option pricing*, Bell J. Econ. 4 (1973), 141–183.

[20] C. Villani, *Hypocoercivity*, Mem. Amer. Math. Soc. 202 (2009).

[21] S. R. S. Varadhan, *Large Deviations and Applications*, SIAM, 1984.

[22] M. Freidlin and A. Wentzell, *Random Perturbations of Dynamical Systems*, 3rd ed., Springer, 2012.